EDGEWOOD

RESEARCH, DEVELOPMENT & ENGINEERING CENTER

U.S. ARMY SOLDIER AND BIOLOGICAL CHEMICAL COMMAND

ERDEC-CR-259

ADVANCED SIGNAL PROCESSING AND PATTERN RECOGNITION METHODS FOR PASSIVE INFRARED REMOTE SENSORS

G.W. Small

OHIO UNIVERSITY
Athens, OH 45701

October 1998

19981120 025

Approved for public release; distribution is unlimited.

Reproduced From

Best Available Copy

Aberdeen Proving Ground, MD 21010-5424

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE 1998 October	3. REPORT TYPE AND DATES COVERED Final; 93 Apr – 97 Aug

4. TITLE AND SUBTITLE

Advanced Signal Processing and Pattern Recognition Methods for Passive Infrared Remote Sensors

5. FUNDING NUMBERS

C-DAAA15-93-K-0003

6. AUTHOR(S)

Small, G.W.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Ohio University, Department of Chemistry, Clippinger Laboratory, Athens, OH 45701

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDEC-CR-259

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

DIR, ERDEC, ATTN: SCBRD-RT, APG, MD 21010-5424

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

COR: Robert T. Kroutil, SCBRD-RT, 410-671-1709

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report describes studies directed to the automated analysis of Fourier transform infrared remote sensing interferogram data. The research presented here demonstrates that both qualitative and quantitative information can be extracted from short segments of digitally filtered interferograms without the need for any background or reference measurement. Through the use of experimental design techniques, an optimization protocol is devised for determining the key implementation parameters of the interferogram-based analysis. An automated compound identification algorithm is developed for trichloroethylene (TCE) and is shown to operate effectively in the presence of a wide variety of infrared backgrounds. A quantitative analysis for sulfur dioxide (SO_2) is also implemented through the direct use of short interferogram segments.

14. SUBJECT TERMS

Interferogram
Digital filtering
Fourier transform infrared spectrometry

15. NUMBER OF PAGES

98

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

Blank

PREFACE

The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for public release. Registered users should request additional copies from the Defense Technical Information Center; unregistered users should direct such requests to the National Technical Information Service.

Acknowledgments

Blank

CONTENTS

FIGURES

TABLES

Blank

ADVANCED SIGNAL PROCESSING AND PATTERN RECOGNITION METHODS FOR PASSIVE INFRARED REMOTE SENSORS

Introduction

Open path Fourier transform infrared (FTIR) spectroscopy is a technique of growing importance in a variety of environmental monitoring applications [1, 2]. In this experiment, an interferometer-based optical system is used to monitor the atmosphere between the spectrometer and an infrared source. Three basic experimental setups are commonly employed, termed the passive terrestrial, active bistatic, and active monostaic configurations. The passive measurement is based on the collection of the naturally occurring infrared emission from some terrestrial source, while the active experiments collect the emission from a commercial blackbody infrared source. In either case, the goal of the analysis is to detect the infrared signatures of target compounds present in the intervening atmosphere between the source and spectrometer.

The analysis of data from these experiments is challenging due to the possible presence of many spectral interferents, as well as the problem of significant changes in the infrared background emission. The latter problem is particularly troublesome in the passive terrestrial experiment due to the complete lack of control of the infrared source radiance.

Recent research in our laboratories has focused on the design of data analysis strategies that meet these challenges [3-13]. This work is based on the application of pattern recognition techniques to identify the characteristic signatures of target compounds directly in the interferogram data collected by the spectrometer. To help reject the contributions of spectral interferents and to overcome the problems associated with a changing infrared background, two preprocessing steps are applied to the interferogram data before the pattern recognition analysis is performed. First, the interferogram is windowed to isolate a short segment displaced from the centerburst. This step helps to discriminate against broad background spectral features whose interferogram representations damp rapidly. By selecting a segment remote from the centerburst, the contribution of these background signatures is minimized. Next, a bandpass digital filter is applied to the windowed interferogram segment. The application of the filter serves to suppress in the interferogram those sinusoidal signals corresponding to spectral frequencies lying outside the filter bandpass. By designing the filter to pass only those frequencies associated with an absorption band of the target compound, frequency selectivity is made a part of the direct interferogram analysis. This prevents any overlap or interference from the interferogram signatures of spectral bands located at frequencies outside the filter bandpass. The interferogram-based analysis thus focuses on a narrow band of spectral frequencies, regardless of the complexity of the infrared spectrum of the analyte or the presence of bands from interfering compounds.

Figure 1 provides an illustration of the application of windowing and bandpass filtering to interferogram data. The first column in the figure plots sections (1300-700 cm^{-1}) of four gas-phase, single-beam FTIR spectra corresponding to a mixture of SF_6 and CCl_3F (A), pure SF_6 (B), pure CCl_3F (C), and an open-beam infrared background (D). Absorption bands are clearly seen in spectra A, B, and C as a decrease in light intensity over the absorbing region. Superimposed on these spectra is a Gaussian-shaped bandpass of a digital filter designed to
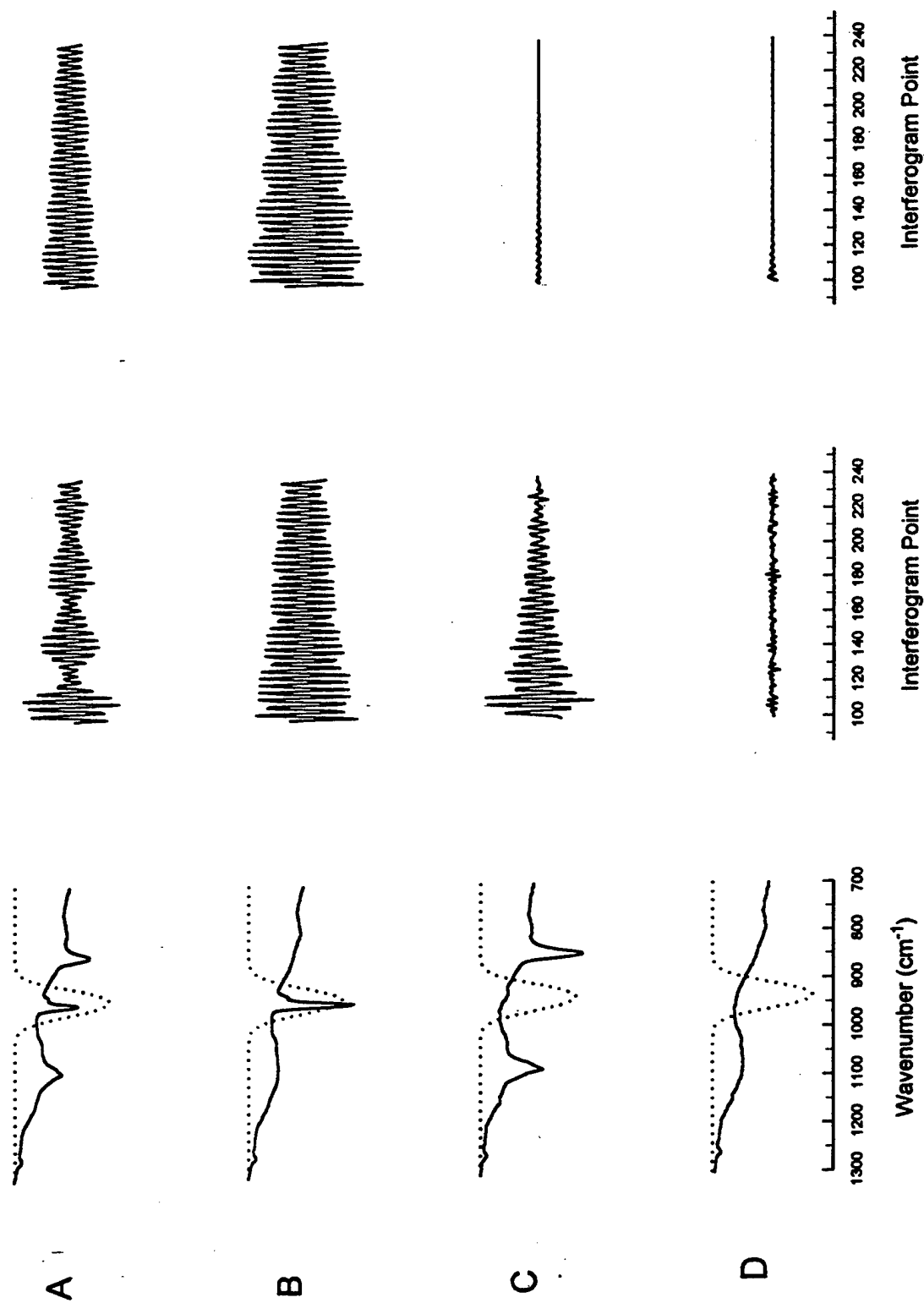
Figure 1. Single-beam FTIR spectra and interferogram segments. Plotted are a mixture of SF_6 and CCl_3F (A), pure SF_6 (B), CCl_3F (C), and an open-beam infrared background (D). Columns 1-3, respectively, correspond to single-beam spectra plotted over 1300-700 cm^{-1}, windowed interferograms plotted over points 100-239, and windowed and filtered interferograms plotted over points 100-239. The interferograms were filtered to isolate the S-F stretching band at 945 cm^{-1}. The filter bandpass used is indicated by the dotted line superimposed on the single-beam spectra.

12

isolate the S-F stretching band of SF_6 at 945 cm^{-1}. The second column of the figure presents the results of windowing the corresponding interferograms to isolate points 100-239 (relative to the centerburst), while the third column plots the same interferogram segment after application of the filter based on the bandpass depicted in the first column. The vertical scale is constant within each column in Figure 1, but differs across the rows.

Inspection of the windowed interferogram segments in the second column reveals the effectiveness of the windowing procedure in removing the contribution of spectral features based on their band widths. As expected, interferogram segment D has a much smaller amplitude than segments A, B, or C due to the more rapid damping of the interferogram representation of the broad infrared background signature. However, segments A, B, and C also illustrate that windowing alone is insufficient to isolate compound-specific information in the interferogram, particularly given the similarities in widths of the spectral bands of most organic compounds. Significant amplitude is clearly seen in each segment due to the contributions of all of the narrow-band spectral features present. Segment A is of particular interest, as the prominent beat pattern in the interferogram arises due to the interference among the representations of the three spectial bands.

As depicted in the third column of Figure 1, the key to isolating compound-specific information in the interferogram is the application of a bandpass filter designed to pass only those frequencies corresponding to an analyte band of interest. Through application of the filter whose bandpass is depicted in the first column of the figure, the interferogram segments in the third column are dramatically altered. The segment corresponding to the infrared background (D) is further reduced in amplitude due to the removal of the contribution of narrow-band noise features. Segment C, containing the contributions of the two bands of CCl_3F (845 and 1084 cm^{-1}) is effectively zeroed also, due to the fact that the frequencies corresponding to these two bands have been suppressed in the interferogram through the application of the filter. Significant amplitude remains in segments A and B, as the filter passes the frequencies associated with the SF_6 band at 945 cm^{-1}. However, the beat pattern in segment A is no longer observed because the frequencies corresponding to the two CCl_3F bands have been suppressed. The interference giving rise to the beat pattern thus no longer occurs. Also, after filtering, the greater magnitude of the SF_6 band in spectrum B relative to that in spectrum A is also seen in the interferogram, manifested as a larger amplitude in the filtered segment. This suggests that both qualitative and quantitative information is present in the filtered interferogram.

This report describes four investigations that employ the basic interferogram signal processing strategies outlined above. First, an experimental design protocol is developed for use in optimizing several adjustable parameters associated with the use of this interferogram-based analysis for qualitative identifications of compound signatures. Second, this methodology is applied to the detection of signatures of trichloroethylene (TCE) in a series of laboratory and open-air monitoring experiments. A wide variety of infrared background conditions are employed in this study. Third, a direct quantitative analysis of sulfur dioxide (SO_2) is implemented with the filtered interferogram data. Controlled field data designed to simulate SO_2 stack emissions are used in this study. Finally, a method is described for developing an interferogram-based compound detection algorithm that does not contain instrument-specific information. This algorithm can be developed with data collected with one spectrometer, and then applied to data collected with a second spectrometer. Sulfur hexafluoride and acetone data are employed in the development of this method.

Experimental Design Protocol for the Pattern Recognition
Analysis of Bandpass Filtered Interferograms

The success of the interferogram windowing and filtering preprocessing steps described above is keyed by optimizing the interferogram segment and filter bandpass chosen for a given analyte. The example displayed in Figure 1 corresponded to a filter bandpass and interferogram segment optimized for extracting the interferogram representation of the SF_6 band at 945 cm^{-1}. This optimization requires the selection of optimal values for four experimental variables: (1) filter bandpass location, (2) bandpass width, (3) interferogram segment starting location, and (4) segment size. In previous work performed in this laboratory, optimal or near-optimal values were obtained for these four experimental variables for a wide range of compounds including CCl_3F, [4], CCl_2F_2 [4], benzene [5], nitrobenzene [3], methanol [11], sulfur hexafluoride [4, 9, 10, 12], acetone [12], and methyl ethyl ketone [12]. In each case, an intensive study was needed to determine the optimal settings. It was empirically noted that relationships among the experimental variables existed. However, no attempt was made to study these relationships in detail.

Other workers have studied the importance of choosing the optimal interferogram segment window to obtain analyte information. The original work in this area was performed for the reconstruction of gas chromatograms from gas chromatography/ FTIR (GC/FTIR) interferogram data [14,15]. In that work, the optimal region of the interferogram was found to be a 100-point segment displaced 60 points from the centerburst. Later work by Bjerga and Small employed bandpass digital filters for the reconstruction of GC/FTIR chromatograms [16]. They concluded that after filtering, the optimal region was a 75-point segment located 171 points from the centerburst. Monfre and Brown employed K-matrix regression to obtain quantitative information from FTIR interferograms. The optimal interferogram window was found to start at interferogram point 10 and end at interferogram point 1388, relative to the centerburst [17,18]. In each of the above studies, it was concluded that it is possible to extract useful analyte information close to the centerburst region of the interferogram. However, no attempt was made to study the relationships that exist between a bandpass filter and the interferogram segment window.

In the work described here, experimental design techniques are used to study the relationships among the four variables involved in an analysis based on bandpass filtered interferograms. The overall goal of the work is to define an experimental protocol for use in optimizing the settings of these variables. This protocol provides an efficient means for designing an interferogram-based detection scheme for any target analyte.

Experimentation

The FTIR data used in this work consisted of laboratory data collected to simulate conditions found in open-air measurements, as well as actual field data collected during a series of field trials. The laboratory data were used to implement the experimental design study of the data analysis variables, while the field data were used to confirm the results of this study.

Sulfur hexafluoride was used as the test analyte in the collection of both types of data. It is a standard test compound used in open path FTIR studies due to its strong absorptivity in the infrared and low toxicity. The S-F stretching band in the region of 945 cm^{-1}

was used as the targeted spectral band in the digital filtering and pattern recognition work reported here. The full width at half-maximum (fwhm) of this band is approximately 10 cm^{-1}.

The laboratory data collection employed a Honeywell emission spectrometer (Model: XM21). This spectrometer design consisted of a flex-pivot "porch swing" Michelson interferometer and employed a closed-cycle Stirling cooler for maintaining the Hg:Cd:Te detector at 77 °K. The detector spectral response was restricted to the 8-12 μm atmospheric transmission window. The spectrometer was aligned with a 4" x 4" extended blackbody infrared source (Model SR-80, CI Systems, Inc., Agoura, CA). This NIST certified source is accurate to ± 0.03 °C and precise to ± 0.01 °C. The blackbody was used to obtain an adjustable temperature source from ambient to 50 °C, thereby simulating changes in the infrared background radiance that might be encountered in an actual open path measurement with the passive terrestrial spectrometer configuration.

For the data collection, a gas syringe was used to inject SF$_6$ samples into a custom short-path gas cell with low density polyethylene windows (0.0005" thickness) [19]. The gas cell was used at atmospheric pressure. The cell body was 8.3 cm long and 16.5 cm in diameter. The cell contained a DC motor driven fan to ensure that a homogeneous mixture of air and SF$_6$ was present throughout the cell [20]. The cell was used at ambient temperature. The actual cell temperature was monitored to ± 0.1 °C with a thermistor probe (Jenco Model 7002H, probe 409B G98598, Jenco Instruments, Inc., San Diego, CA). Over the course of the data collection, the cell temperature varied from 24.2 - 25.9 °C.

The cell was positioned between the blackbody source and spectrometer, with a distance of 10.8 cm between the source and cell and 14.6 cm between the cell and spectrometer. A helium neon laser was used to align the blackbody source, cell, and spectrometer such that the spectrometer field of view contained only the cell and the source.

Interferograms were collected with two different volumes of SF$_6$ (0.1 and 0.2 cm^3) and several blackbody temperatures. The SF$_6$ gas volumes correspond to concentrations of 56.3 and 112.7 ppm, respectively. The corresponding concentration-path length products were 4.7 ppm-m and 9.4 ppm-m. In addition, interferograms were collected at each source temperature with no SF$_6$ in the cell and with no cell in the optical path. Table 1 summarizes the data collected. All interferograms were single scans (i.e., no signal averaging was performed) consisting of 1024 points sampled at every eighth zero-crossing of the reference laser. The maximum observable frequency was 1974.8 cm^{-1} and the point spacing in the transformed spectra was 3.9 cm^{-1}.

The occurrence of small temperature differences between SF$_6$ in the gas cell and the blackbody source produced some cases in which even though SF$_6$ was present in the cell, its spectral band at 945 cm^{-1} could not be detected visually. These data were retained, however, and the interferograms were still assigned to the SF$_6$-containing data class. It is estimated that these interferograms account for approximately 1-2% of the data.

The collection of the field data spanned a period of eight weeks and employed a portable emission spectrometer constructed by Midac Corp. (Irvine, CA). The spectrometer employed a linear drive Michelson interferometer and a 1 mm^2 liquid nitrogen cooled Hg:Cd:Te detector. Single-scan interferograms were collected with the same characteristics as described above.

Table 1

Laboratory Data

Sample volume (cm^3)	Blackbody Temperature (°C)	Cell Temperature (°C)	Number of interferograms	
			SF$_6$-containing	Background
0.2	Ambient	24.6	48	64
	20.0	24.7	32	64
	22.0	24.8	32	64
	24.0	24.9	32	64
	24.9	25.0	32	64
	25.0	25.1	32	64
	26.0	25.2	32	64
	28.0	25.2	32	64
	30.0	25.3	32	64
	35.0	25.3	32	64
	40.0	25.4	32	64
	45.0	25.4	32	64
	50.0	25.2	32	64
0.1	Ambient	24.6	64	64
	20.0	24.8	32	64
	22.0	25.1	32	64
	23.0	25.2	32	64
	24.0	25.3	32	64
	24.5	25.5	32	64
	25.0	25.5	32	64
	25.5	25.6	32	64
	26.0	25.7	32	64
	27.0	25.7	32	64
	28.0	25.8	32	64
	30.0	25.8	32	64
	35.0	25.8	32	64
Total			880	1664

The data collection employed two different implementations of the passive terrestrial spectrometer configuration. First, the spectrometer was mounted on a tripod and used to view a variety of terrain backgrounds, both with and without SF_6 being released in the field of view. Second, the spectrometer was mounted in a shock-absorbing assembly and placed in a helicopter with the field of view of the spectrometer being directed at the ground. The helicopter made aerial passes past a ground source of SF_6.

A total of 40,344 interferograms were collected in these field experiments. This data set was reduced to 4000 interferograms (2000 SF_6-containing, 2000 background) through the application of a data set selection algorithm reported by Carpenter and Small [7]. The set of 4000 interferograms was further subdivided randomly into a training set of 3000 interferograms for use in developing the digital filtering and pattern recognition methodology and a separate prediction set of 1000 interferograms used for testing. The SF_6-containing and background interferograms were selected separately in order to maintain equal class sizes in both the training and prediction sets. Table 2 describes these data sets. The determination of whether or not an interferogram contained SF_6 information was made by Fourier transforming the interferogram to the spectral domain, subtracting a background spectrum and visually inspecting the resulting difference spectrum for the presence of the S-F band at 945 cm^{-1}. Through the application of this procedure, each of the 4000 interferograms was judged either an SF_6-containing or a background interferogram. This assignment procedure is inexact when working with field data due to the changing infrared background emission in the passive terrestrial experiment. The difficulty in matching each spectrum to an appropriate background spectrum results in a variety of artifacts in the difference spectra that can obscure weak analyte signals. Nevertheless, we estimate that this visual inspection procedure has an assignment error rate no greater than 3-5%. Due to the physical movement of the sample into and out of the optical path over time, however, there is no better assignment procedure available.

For the data analysis, the collected interferograms were transferred to a Silicon Graphics 4D/460 computer operating under the Irix operating system (version 4.0.5, Silicon Graphics, Inc., Mountain View, CA). The digital filtering and pattern recognition calculations reported here were performed on this system with original software written in FORTRAN 77 and C. Analysis of variance computations, the calculation of normal scores, and the construction of the main and interaction effects plots were performed with the Minitab statistical software package (version 10, Minitab, Inc., State College, PA) implemented on a Dell 466/L computer operating under Microsoft Windows (version 3.1) and MS-DOS (version 6.2, Microsoft, Inc., Redmond, WA).

Results and Discussion

Overview of Interferogram Analysis Methodology. The interferogram analysis techniques used in this work employed digital filtering and pattern recognition methods. The digital filtering method used implements a time-varying finite impulse response filter. This filter design technique employs interferogram data in the calculation of the filter and has been found to perform well in comparison to other filter design schemes [8]. The time-varying nature of the filter helps to match the filter to the rapidly damping exponential character of the interferogram signal. These filters have the form

$$y[n] = \sum_{i=1}^{f_n} h_n[i] \, x[n-o_n[i]] \qquad\qquad (1)$$

where y[n], the intensity of filtered interferogram point n, is computed from a convolution sum of f_n terms. The summation is based on the products of an impulse response function, $h_n[i]$, and the intensities of selected points in the unfiltered interferogram, $x[n-o_n[i]]$. The points used in the unfiltered interferogram are specified relative to point n. The time-varying nature of this filter is achieved by having separate $h_n[i]$, $o_n[i]$, and f_n for each n. The design of the filter requires the specification of the filter frequency response function and a set of interferograms to use in the computation of the $h_n[i]$ and the selection of f_n and the $o_n[i]$. The frequency response is held constant with time, but the optimal implementation of the filter in the time domain is allowed to vary with interferogram point (i.e., with time). For a given error level in the bandpass approximation achieved by the filter, this scheme allows filters to be generated with fewer coefficients than would be required with a fixed coefficient filter [8].

For the work reported here, separate filters were generated for the laboratory and open path data. The entire set of 2544 interferograms was used in the calculation of filters for the laboratory data, while the training set of 3000 interferograms was used with the open path data. These interferograms define the number of observations used in a multiple linear regression calculation of the $h_n[i]$ [8].

The frequency response and impulse response are Fourier transform pairs. The functional form (i.e., shape), bandpass width, and bandpass position of the frequency response are user-specified variables. The filter design computation attempts to achieve the desired shape, width, and position of the frequency response in a filter that can be applied directly to the interferogram through the use of eqn. 1. For this work, the bandpass shape was Gaussian, and the bandpass position and width constituted two of the variables to be explored [8]. Bandpass shapes other than Gaussian can be used although our previous work indicates that this variable is much less significant than either the bandpass position or width.

Once the interferogram has been filtered, recognition of the signature of a target compound is achieved through the application of pattern recognition techniques to the filtered interferogram segment. Pattern recognition methods treat an m-point interferogram segment or "pattern" as a vector in an m-dimensional space. Recognition of the signature of a target compound is based on clustering in the m-dimensional space of the points representing the filtered interferogram segments. If these points are clustered in a manner that allows them to be discriminated based on the presence of the target compound, pattern recognition techniques can be used to implement an automated procedure for estimating compound presence, given a filtered interferogram segment.

The pattern recognition technique employed in this work was piecewise linear discriminant analysis (PLDA) [9]. PLDA is one of a number of general pattern recognition algorithms for use in classifying data vectors into two or more categories. It offers the advantage of handling nonlinear relationships among the data vectors while being computationally fast enough to be compatible with large data sets. In this regard, it offers several advantages over competing methods such as artificial neural networks. Through the use of a representative "training set" of data, PLDA computes the position of a set of linear surfaces in the data space in

18

Table 2

Open Path Data

Type of measurement	Number of interferograms			
	Training set		Prediction set	
	SF_6-containing	Background	SF_6-containing	Background
Stationary	1206	767	409	247
Airborne	294	733	91	253
Total	1500	1500	500	500

an effort to define boundaries between patterns belonging to different data categories or classes. In the remote sensing application, two data classes exist, corresponding to the presence (Class 1) or absence (Class 2) of a target compound.

Each linear surface is defined by the locus of points lying orthogonal to an optimally positioned unit vector termed a weight vector or linear discriminant. The piecewise linear discriminant is defined by the set of individual weight vectors that together form a piecewise linear approximation to a nonlinear separating surface between the data classes. In the work described here, these vectors were computed in a stepwise manner. The first weight vector was positioned in an optimal orientation, followed by positioning of the second vector to form an optimal two-vector piecewise linear discriminant. Thus, calculation of the pth weight vector was based on the positioning of a vector to combine with the p-1 vectors previously computed to form a p-vector discriminant.

One of the requirements for the piecewise linear discriminant is that each weight vector is "single-sided". This means that the vector defines a linear surface that partitions the data space such that one side of the surface contains members of only one data class. This "pure" side of the surface is distinguished from the other ("mixed") side which can contain members of all other data classes. In applying PLDA to interferogram analysis, we have established the convention that the pure side of the separating surface corresponds to compound-containing interferograms.

Each weight vector is positioned through the use of numerical optimization techniques. The optimization seeks the optimal value of a response function which encodes the ability of the discriminant to classify patterns correctly based on their known class identities [13]. In addition, the response function penalizes weight vectors that are not single-sided. In our work, Simplex optimization was used to position the weight vectors. Applied to PLDA, Simplex optimization is based on the iterative movement of a set of candidate weight vectors, with each iteration attempting to replace one candidate vector with a new vector that achieves a more optimal value of the response function. In the work described here, the Simplex optimization was operated in a non-interactive manner based on a protocol developed through experience with the technique. This protocol employed a specified discriminant size (i.e., number of weight vectors comprising the piecewise linear discriminant), number of Simplex initializations used in computing each weight vector, and number of iterations performed before reinitializing the optimization. The initialization procedure required the specification of a "spanning constant", a numerical value used to form the initial set of weight vectors through perturbation of a single input vector. The input weight vectors used to start the Simplex optimization were computed directly by use of the Bayes linear discriminant procedure [21].

Once the piecewise linear discriminant is computed, the classification of any filtered interferogram segment, x_i, can be performed as

$$d_i = max (w_1^T x_i, w_2^T x_i, ..., w_p^T x_i) \qquad (2)$$

$$d_i > 0 \mid Class\ 1 \qquad (3)$$
$$\leq 0 \mid Class\ 2$$

20

where d_i is the discriminant score for x_i, computed as the maximum vector dot product formed between x_i and the p individual weight vectors (w_1, w_2, ..., w_p). By our convention, discriminant scores greater than zero signal data points (i.e., filtered interferogram segments) lying on the "compound-present" side of the piecewise linear discriminant.

Description of Filter Design Variables Studied. The design of an optimal filter is a key step in applying the interferogram-based detection methodology described above to a new compound. The limit of detection of an analyte is largely based on the degree to which the values of four filter design variables or factors are optimized. The variables are: (1) filter width, (2) filter position, (3) interferogram segment length, and (4) interferogram segment starting position (relative to the centerburst position).

Figure 2 depicts the first two variables. Shown are the frequency responses of two Gaussian-shaped bandpass digital filters (dashed lines) superimposed on a single-beam infrared spectrum exhibiting the SF_6 absorption at 945 cm^{-1} (solid line). Both filters shown in Figure 2 are positioned centered on the analyte band. The filter is positioned near the analyte band so that those frequencies will pass through the filter.

The width of a filter is an important factor in its effectiveness. As shown in Figure 2, the wider the filter, the more frequencies (both analyte and background) that will be allowed to pass through the filter. An important characteristic of filter design is that decreasing the width of the filter frequency response requires an increase in the number of points in the impulse response ($h_n[i]$ in eqn. 1). The number of computations required to implement the filter is thus related to the filter width.

In the design of an optimal filter, a joint effect on filter effectiveness is expected between filter position and filter width. This can be rationalized by noting that wide filters do not need to be centered directly on the analyte band to pass an equivalent number of analyte frequencies compared to narrow filters centered directly on the band.

Figure 3 depicts the other factors involved in digital filter generation. Shown is an interferogram collected when SF_6 was present in the optical path of the spectrometer. The interferogram segment length determines the amount of information that the pattern recognition analysis technique can use to distinguish between compound-containing and background patterns. The limit of detection is determined by the amount of analyte information present relative to the "noise level" defined by the variation in the background patterns. Although longer interferogram segments generally outperform shorter segments, the use of shorter segments is computationally more efficient. The goal is therefore to use an interferogram segment length as short as possible without negatively affecting the limit of detection.

The interferogram segment starting point is also an important factor. Because the interferogram signal decays exponentially, the closer the interferogram segment is to the centerburst region, the more analyte information will be present. However, more background information will also be present in segments near the centerburst than in segments displaced from the centerburst. This can be rationalized by considering that, after filtering, the interferogram has been reduced to contain the contributions of only two spectral features: the analyte band and the infrared background signature. By applying the filter, the infrared background has been truncated in frequency to coincide with the frequency response of the filter. This can be seen visually in Figure 2 where the overall infrared background is considered

21
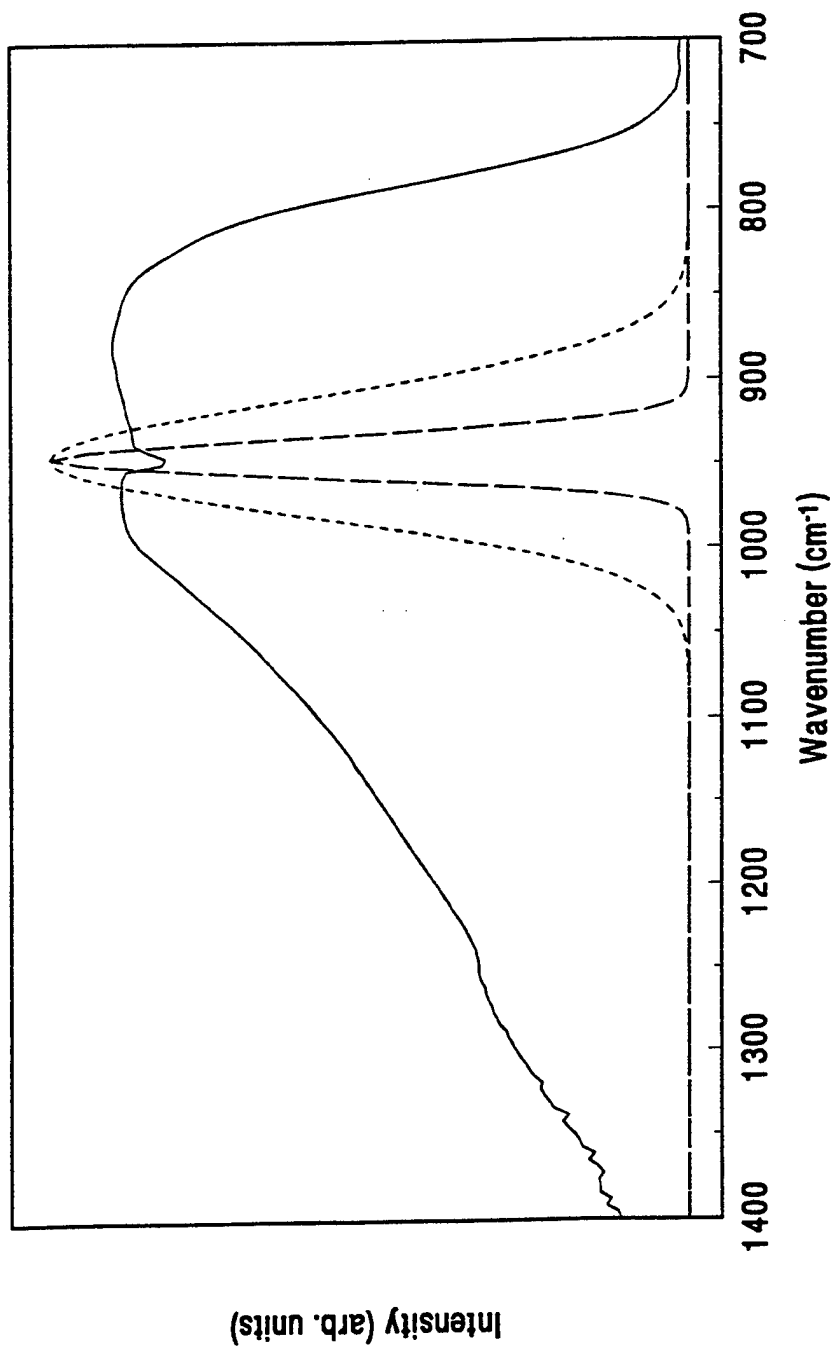
Figure 2. Single beam spectrum of SF$_6$ (solid line) with frequency responses of two Gaussian-shaped bandpass digital filters (dashed lines) superimposed.

22
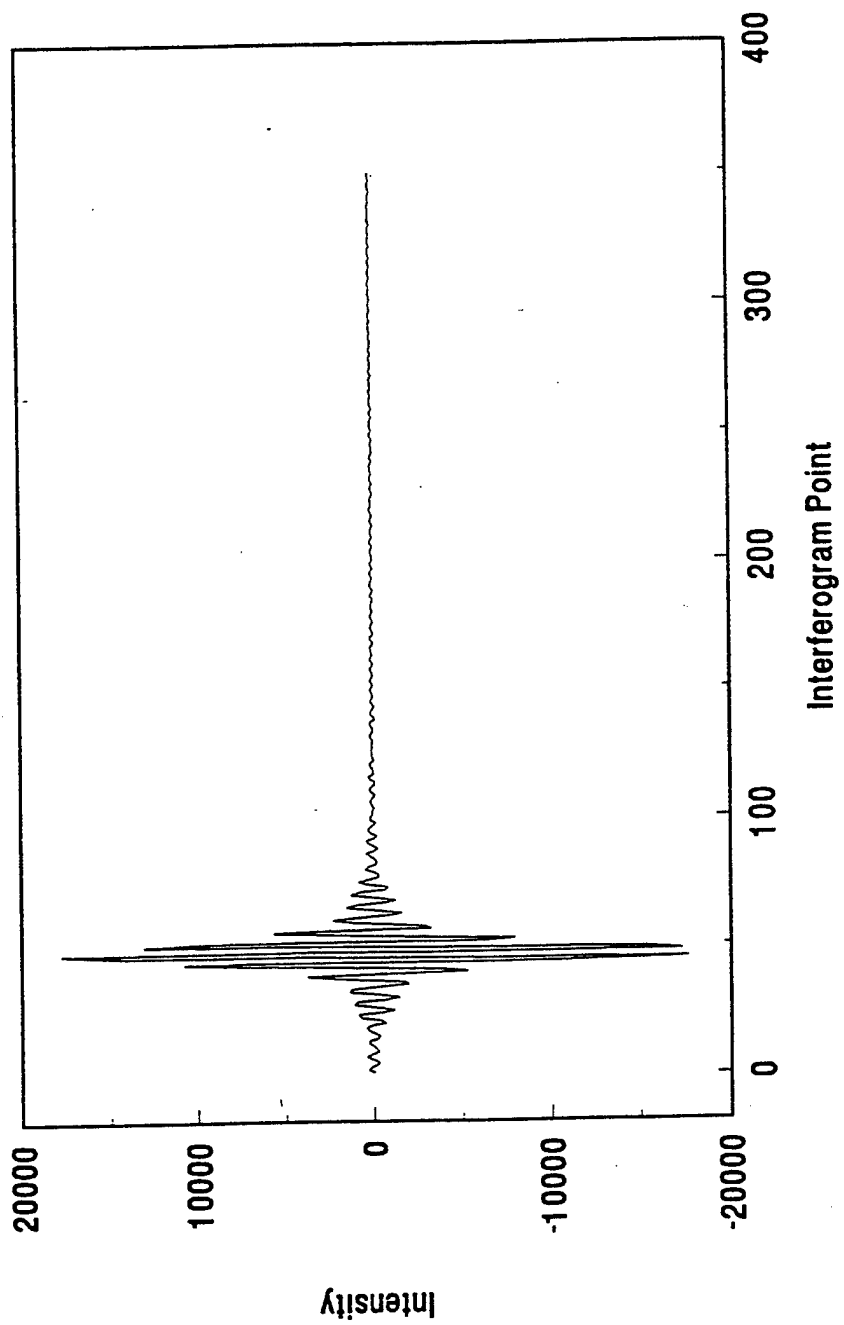
Figure 3. 1024-point FTIR interferogram collected when SF$_6$ was present in the field of view of the spectrometer.

to be the single beam spectrum minus the SF_6 absorption band. Application of either filter depicted in the figure truncates the single beam spectrum to coincide with the frequency response of the filter. The SF_6 band is then superimposed on the frequency response function. The two spectral features remaining after application of the filter differ in width. The corresponding representations of these features in the interferogram also differ in width, with the representation of the narrow analyte spectral feature decaying more slowly than that of the wider frequency response function. This suggests that some optimal segment starting point exists in the interferogram that balances the overall decay of the analyte signal vs. the difference in rates of decay between the analyte and background signals.

It can be argued that a joint effect on filter effectiveness must exist between interferogram segment length and starting position. The effect of having less analyte information for the pattern recognition analysis in short segments can be partially overcome by judiciously choosing a segment starting position closer to the centerburst. Conversely, longer segments may still have adequate analyte information for the pattern recognition analysis using interferogram segments distant from the centerburst. The above discussion suggests that an additional joint effect which must be considered is the relationship between interferogram segment starting position and filter width. These variables are correlated due to the change in the rate of decay of the interferogram signal with filter width. Thus, the use of narrow filters in conjunction with interferogram segments near the centerburst will include more background information than would be included if the same segment were used with a wider filter.

Optimization of the four filter design variables involves setting discrete values or levels for each factor, followed by the generation and testing of filters based on the selected values. The key to computational efficiency in the optimization lies in minimizing the number of filters that must be generated and tested. This must be done judiciously, however, as the relationships among the factors determine the degree to which the optimal value of one factor depends on the value of another factor.

The above discussion illustrates that, on theoretical grounds, several pairwise relationships must exist among the four factors. Other less obvious two-way, three-way, and higher order relationships among the factors may also exist. The computational effort in the optimization must be made where it will provide the most benefit and where the strongest relationships among the factors exist. For example, an extensive joint study of two factors is of little real value if those factors are not strongly related. Optimization of these factors could be performed independently, thereby eliminating the need for a joint study in which the values of both factors are studied together. Thus, knowledge of the significance of each factor and the relationships among the factors is critical in devising a protocol for the optimization that will: (1) allocate the greatest resources to the optimization of the variables that are most significant in influencing the limit of detection; (2) lead to an overall optimal or near-optimal filter design; and (3) minimize the computational requirements of the optimization. The goal of this work is to establish such a protocol for the filter optimization by formally studying the significance of the relationships among the four filter design variables.

Experimental Design. To study the relationships among the four variables and to find the optimal variable settings, a formalized statistical experimental design was performed [22-26]. Experimental designs allow the examination of the main effects of the experimental variables, as well as the joint or interaction effects among the variables. One main effect is defined for each variable, encoding the effect on the experimental result of changing the settings of the variable.

24

By contrast, an interaction effect is defined for each combination of variables. These effects encode the influence on the experimental result of making joint changes in the variable settings. To ensure complete exploration of these main and interaction effects, all possible combinations of the different factor levels must be examined. This type of experimental design is termed a full factorial design.

In a full factorial experimental design, a response function must be used that describes the overall performance or the effectiveness of the variable settings. Such a function has the form

$$R = f(p_1, p_2, ..., p_n) \qquad (4)$$

where R is the value of the response function for given settings of the n variables, p_i. The response function numerically encodes the degree to which the variable settings produce an optimal result. In the present application, an optimal result is the lowest possible limit of detection of the target analyte. Together with the factor settings, the response function defines a response surface whose shape dictates the manner in which the individual variables and their joint effects impact on the optimal limit of detection. Each main and interaction effect is expressed in the units of the response function.

The response function chosen for this study was the actual pattern recognition detection performance observed from application of PLDA to the filtered interferogram data. The filter used in each case corresponded to specific settings of the four filter design variables. Through the use of the training set data in which the classifications (SF_6-containing or background) were known, the classification performance achieved by the pattern recognition analysis was chosen to serve as an indicator of the degree to which the factor settings were optimal.

A full factorial design study was performed based on the variables and levels specified in Table 3. Five filter bandpass widths, five filter positions centered around the SF_6 band at 945 cm^{-1}, five interferogram segment starting points relative to the centerburst, and five interferogram segment lengths were studied. The choice of levels for each variable was based on previous experience in applying the interferogram-based methodology to the detection of SF_6. The use of filter widths significantly wider than the SF_6 band width (fwhm ≈ 10 cm^{-1}) is based on our experience that the inclusion of some background information helps the pattern recognition methodology discriminate between analyte-containing and non-analyte interferograms. In addition, the filter generation procedure used here tends to produce filters with poor attenuation characteristics when a very narrow filter bandpass is specified (e.g., < 30 cm^{-1} fwhm).

In a full factorial design study based on four variables and five levels for each, a total of 5 x 5 x 5 x 5 = 625 possible combinations of the variables are possible. For each variable combination, two replicate training procedures were performed resulting in 1250 evaluations of the response function. Replication was performed by changing the signs of the spanning constants used in the training protocol. This causes the Simplex optimization to be initialized differently and thus results in a different final discriminant. The purpose of replication is to obtain a value for the experimental error present in the iterative training procedure. The experimental error will be employed in the statistical analysis of the results.

For each set of factor settings, a digital filter was generated, and a four-vector piecewise linear discriminant was computed. The value of R in eqn. 4 was defined as the number of SF_6-

Table 3

Variables and Levels Used in Factorial Design

Variable	Levels
[a]Filter bandpass width (wd)	36.4, 45.4, 54.5, 63.6, 72.7 cm^{-1}
Filter bandpass position (fp)	937.0, 940.9, 944.7, 948.6, 952.4 cm^{-1}
Interferogram segment length (sl)	60, 80, 100, 120, 140 points
[b]Interferogram segment location (sp)	Starting point 75, 100, 125, 150, 175

[a]Full width at half-maximum (fwhm).

[b]Relative to interferogram centerburst.

containing patterns correctly classified by the four-vector discriminant. Background interferograms were not considered in computing the response function, as the single-sided requirement of the piecewise linear discriminant dictates that no background interferograms in the training set will be misclassified.

As indicated in Table 1, the total number of SF_6-containing patterns in the training set was 880. The set of factor settings which classified the most interferograms correctly consisted of interferogram segment starting point 100, segment length 140 points, filter position 937.0 cm^{-1}, and filter width 45.4 cm^{-1} (fwhm). This set classified 860 of the 880 patterns correctly (97.7 %). The factor setting which classified the least number of patterns correctly was interferogram segment starting point 75, segment length 60 points, filter position 937.0 cm^{-1}, and filter width 36.4 cm^{-1} (fwhm). This set classified 671 active patterns correctly (76.3 %). The mean and standard deviation of the number of patterns correctly classified in the 1250 cases were 803.5 and 19.5 patterns, respectively. Figure 4 is a histogram showing the distribution of the number of patterns correctly classified.

Analysis of Variance. Analysis of variance (ANOVA) techniques [22-26] were used to estimate the main and interaction effects from the PLDA results. In ANOVA, response function values corresponding to the variable settings are fit to a least-squares model that separates the variance in the response function into assignable causes (i.e., main and interaction effects) and random variation. The model employed in this study was a fixed effect model of the form

$$Y = \mu + sl + sp + fp + wd + sl \times sp + sl \times fp + sl \times wd + sp \times fp + sp \times wd \qquad (5)$$

$$+ fp \times wd + sl \times sp \times fp + sl \times sp \times wd + sl \times fp \times wd + sp \times fp \times wd + sl \times sp \times fp \times wd + e$$

where Y is the response function value for a specific combination of variable settings, μ is the overall mean value of the response function and e is the error term that estimates the random variation. The other terms in the model are the main and interaction effects based on the combinations of interferogram segment length (sl), segment starting position (sp), filter bandpass position (fp), and bandpass width (wd). For example, sl×sp is the two-way interaction effect between segment length and segment starting position.

The results from the ANOVA study are shown in Table 4. The first column in the table is the source of the variance (i.e., main and interaction terms), while the second column lists the degrees of freedom corresponding to each term. The third and fourth columns are the sum of squares and mean square, respectively, for each main and interaction term. The significance of each main and interaction effect can be evaluated by performing F-tests to compare each mean square to the mean squared error. The computed F-values are listed in the fifth column in Table 4. The F-values can be employed to compute a probability that a given effect is significant. The null hypothesis associated with each probability is that, within sampling variability encoded in the degrees of freedom, the corresponding mean square is equal to the mean squared error. Probabilities near zero indicate that the null hypothesis can be rejected with a high degree of confidence (i.e., the effect is significant).

The interpretation of the results in Table 4 is not straightforward. The probabilities computed from the F-values indicate that all main and interaction effects are statistically significant (probability < 0.05). This result is not surprising due to the small error in the iterative training procedure employed in the pattern recognition. This error is computed as the square root of the mean squared error in Table 4 ($[59.0]^{1/2} = 7.7$ patterns).

27
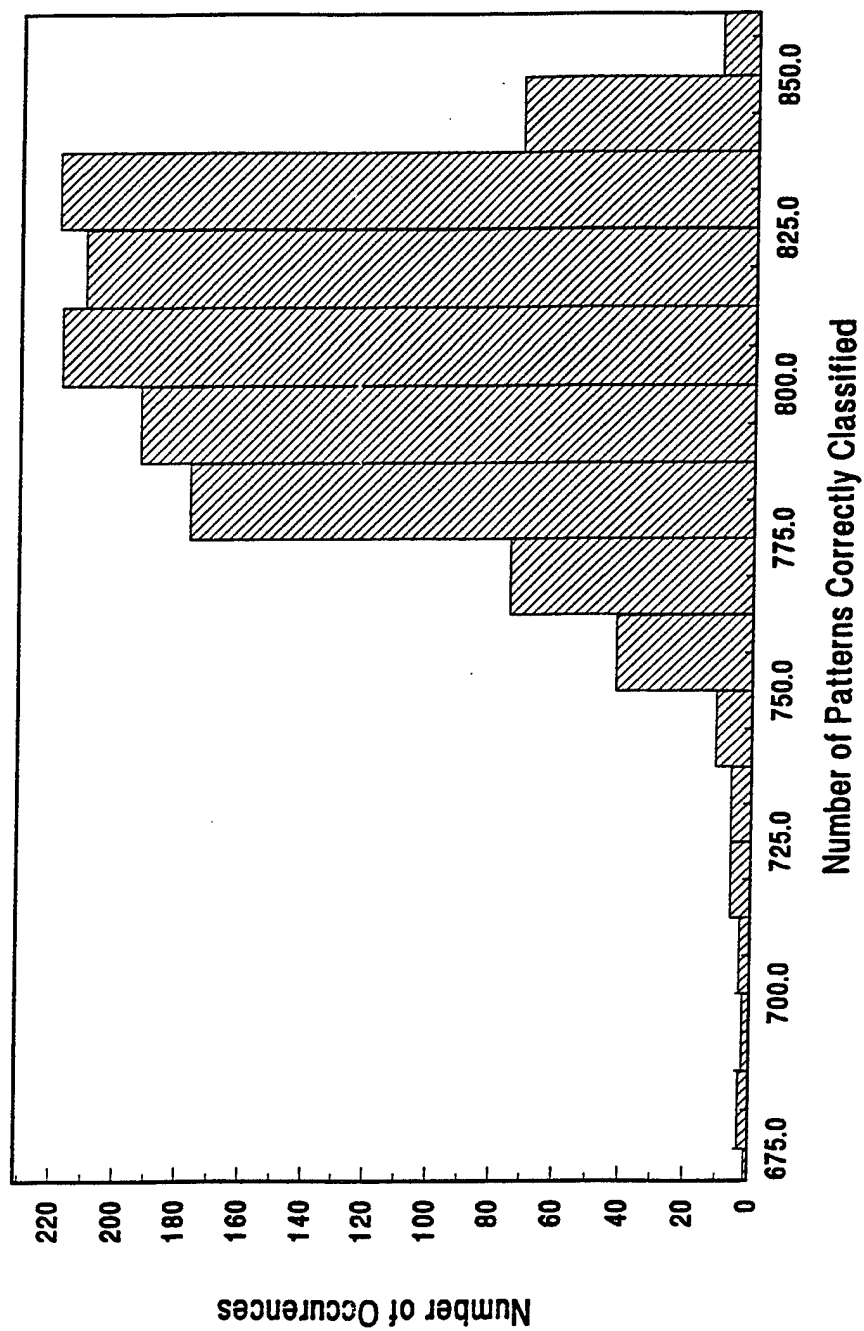
Figure 4. Histogram displaying the distribution of the pattern recognition results from the 1250 sets of variable settings.

Table 4

Analysis of Variance Table

Source	Degrees of Freedom	Sum of Squares	Mean Square	F
sl	4	475306.4	118826.6	2014.2
sp	4	88174.0	22043.5	373.7
fp	4	5553.5	1388.4	23.5
wd	3	10987.8	2747.0	46.6
sl×sp	16	46980.1	2936.3	49.8
sl×fp	16	5080.0	317.5	5.4
sl×wd	16	1671.9	104.5	1.8
sp×fp	16	4366.5	272.9	4.6
sp×wd	16	53382.6	3336.4	56.6
fp×wd	16	18700.2	1168.8	19.8
sl×sp×fp	64	7924.5	123.8	2.1
sl×sp×wd	64	18874.9	294.9	5.0
sl×fp×wd	64	11453.3	179.0	3.0
sp×fp×wd	64	45693.0	714.0	12.1
sl×sp×fp×wd	256	39623.4	154.8	2.6
Error	625	36872.0	59.0	
Total	1249	870644.1		

Some of the effects are clearly more significant than others based on their *F*-values. Using the relative size of the F-value as the criterion for the significance of an effect in influencing pattern recognition performance, all main effects are again judged significant. The two-way interaction terms that were expected to be influential (i.e., slxsp, spxwd, and fpxwd) are also found statistically significant by this criterion. The other three two-way interactions, as well as three of the four three-way interactions and the four-way interaction are judged not significant based on their lower *F*-values. The only significant three-way interaction was spxfpxwd. The importance of this interaction confirms that an adjustment of filter position impacts the optimal filter width and that optimization of these two parameters must consider the segment starting position.

The conclusions drawn from the ANOVA results displayed in Table 4 are based on the assumptions that the model described by eqn. 5 is appropriate and that the model residuals are drawn from a normal distribution. The appropriateness of the model determines whether the error estimate is valid. This is an important consideration, given that the *F*-values are based on the computed mean squared error. Furthermore, correct interpretation of the computed *F*-values is based on the assumption of normality.

To evaluate these issues, the residuals from the least-squares fit of eqn. 5 were studied graphically. Figure 5 is a plot of the 1250 residuals vs. the corresponding fitted values predicted by the model. Although the linear correlation coefficient is 0.00, the cone-shaped appearance of the plot suggests the presence of nonconstant variance.

The issue of normality was addressed by constructing a normal probability plot of the residuals. Blom's method [27] was used to estimate the residuals that would be obtained from a normal distribution. A linear or near-linear relationship between the actual and estimated residuals provides evidence that the residuals are normally distributed. Figure 6 plots the actual vs. estimated residuals. This plot exhibits a sigmoidal shape with a linear correlation coefficient of 0.924. For n = 1250, this value of the correlation coefficient is not statistically significant.

One approach to overcoming the lack of a normal distribution and nonconstant variance in the model residuals is to perform a suitable transformation of the response variable. The purpose of the transformation is to compute response values which are more normally distributed. One general transformation approach is the Box-Cox transformation [24,28]. After applying a suitable Box-Cox transformation to the response values, the ANOVA calculations were repeated. The results showed that after transformation, most of the nonconstant variance was removed. The residuals were also more normally distributed. A plot of actual vs. estimated residuals yielded a linear correlation coefficient of 0.967. However, the *F*-values for each of the effects were nearly the same as before transformation. Thus, making the response values more normally distributed by transformation had little effect on the conclusions from the ANOVA study.

While performing the Box-Cox transformation, an interesting observation was made. The presence of nonconstant variance and the lack of a statistically significant normal distribution was found to be due to the large interaction effect present between interferogram segment starting point and filter width (spxwd). From the histogram shown in Figure 4, it was found that response function values from variable combinations including interferogram starting point 75

30
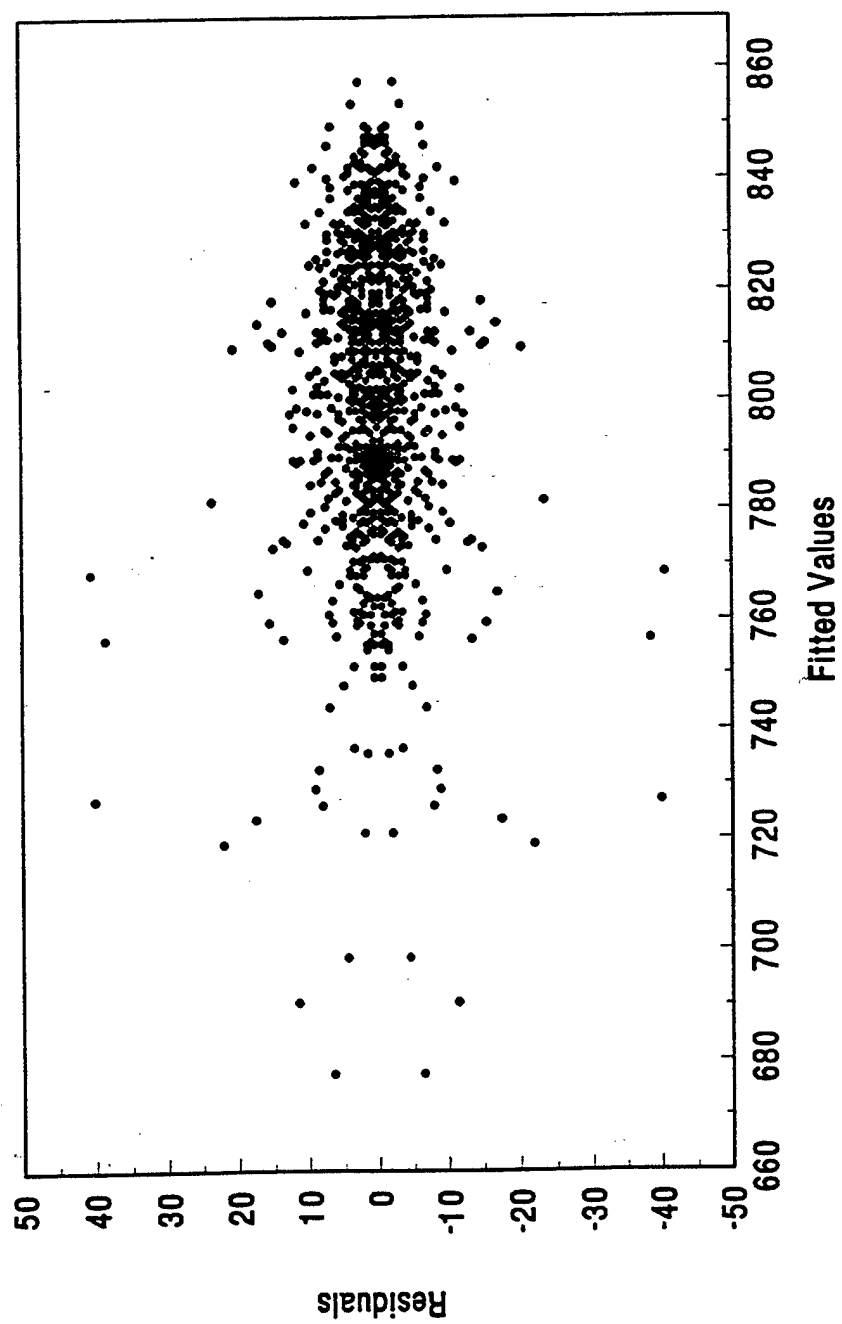
Figure 5. Plot of residuals from the least-squares fit of the ANOVA model shown in eqn. 5 vs. the corresponding fitted values predicted by the model.
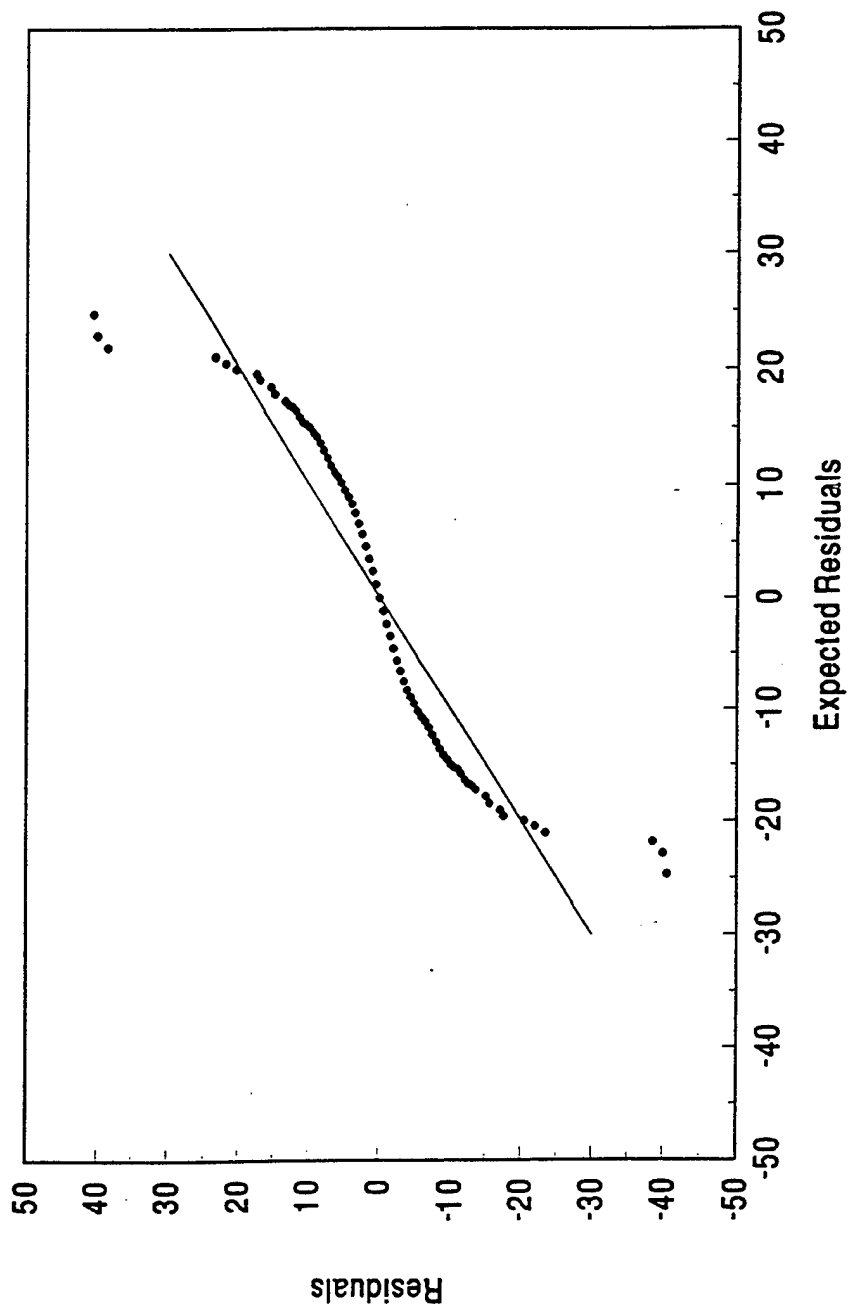
Figure 6. Normal probability plot of residuals from the least-squares fit of the ANOVA model shown in eqn. 5 vs. the estimated residuals. The plot has a correlation coefficient of 0.924.

comprised the lower tail of the histogram. Thus, the ANOVA model (eqn. 5) could not adequately account for this large effect. This hypothesis was tested by removing those response function values and using ANOVA to reestimate the main and interaction effects. The model residuals were now found to contain constant variance and the normal probability plot of the residuals showed only a slightly sigmoidal shape with a linear correlation coefficient of 0.990. These results suggest that removing the response values corresponding to filters computed with interferogram starting point 75 would be a viable approach. However, we do not feel such a drastic measure is justified. The conclusions drawn from this ANOVA study are based on the relative magnitudes of the F-values and not their associated probabilities. This approach produces a clear distinction between the significant and insignificant effects. In this case, the presence of nonconstant variance and a nonnormal distribution do not negatively impact the study.

Protocol for Filter Design. Using the conclusions drawn from the ANOVA results, a protocol for the filter generation was developed. Since the interactions among interferogram segment starting position, bandpass width, and bandpass position are highly significant, these variables must be optimized together. Based on the computed F-values in Table 4, interferogram segment length and starting point were shown to have the greatest influence on filter performance. Further evidence for this conclusion is observed in Figure 7. This figure displays a series of plots showing the mean response function score for each variable setting (solid lines) and the overall mean response function score for all 1250 experiments (dashed line). The variable settings on the horizontal axis are in numerical order identical to Table 3. For example, the left plot shows the mean response function score for the five variable settings for interferogram segment length. Within this plot, the leftmost variable setting indicated by the tick on the horizontal axis corresponds to a segment length of 60, while the rightmost variable setting corresponds to a segment length of 140.

Figure 7 can be used to estimate the main effects for each variable setting. The main effect for a given variable level is computed as the mean response function score for all treatments involving that level minus the overall mean response function score. The main effects are employed to make comparisons about the effect of a variable setting on the response. To interpret main effects properly, it is important to remember that the estimated error of the pattern recognition training procedure was found to be approximately 8 patterns and should be employed as a reference for any comparisons.

Based on the magnitudes of the main effects, interferogram segment length and starting point had the most influence on the pattern recognition performance. As expected, there is a steady improvement in the pattern recognition performance as the interferogram segment length increases. The difference between the main effects for the longest and shortest interferogram segment lengths was 55 patterns. This represented the largest difference in main effects between two settings for any of the four variables.

Due to the large main effects, it is also evident that segment starting position is the second most influential parameter. There is a steady improvement in performance as the interferogram segment approaches the centerburst region of the interferogram (i.e., interferogram segment starting point approaches zero). However, the main effect for the closest segment starting point (75) is less than for starting point 100. This suggests that the majority of the SF_6 signal is compressed into a short region of the interferogram located after the bandpass filter signal has damped to zero. For the narrow bandpass filters, the bandpass filter information
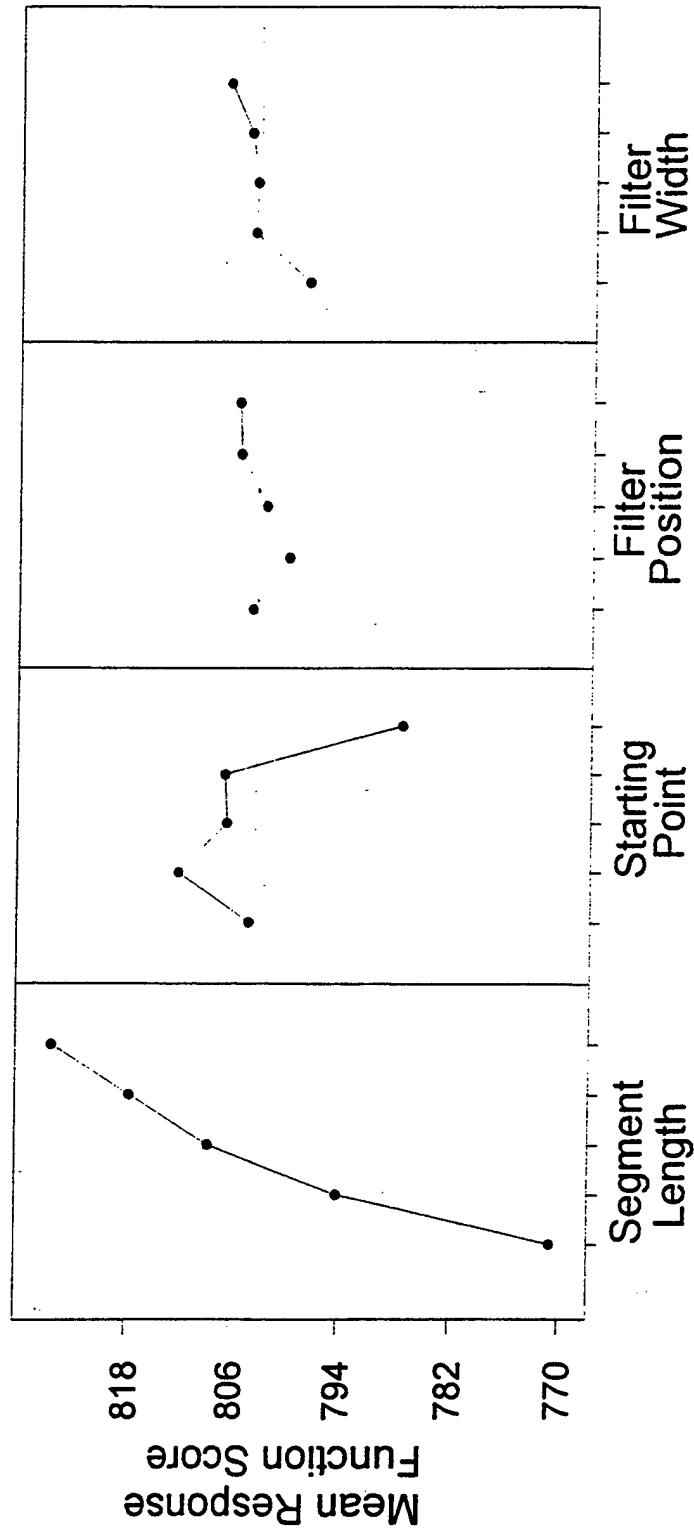
Figure 7. Main effects plots displaying the mean response function score for each variable setting (solid line). The dashed line indicates the overall mean response function score. The settings on the horizontal axis correspond to each level for that variable. The vertical axis is in units of the response function.

does not appear to damp to zero until the region near point 100 in the interferogram. Thus, the narrow filters based on interferogram segments near the centerburst will perform poorly and wide filters based on the same segment will perform well. This association between segment starting position 75 and filter width is the cause of the large F-value present for the spxwd interaction term in the ANOVA model.

The main effects for each level of filter position and width do not follow such a discernible pattern, however. The main effect for each filter position setting increases as the bandpass is moved to higher frequencies. However, the difference in the magnitudes between the largest and the smallest main effects is small (~6) indicating that filter position is less influential than segment length or starting position on pattern recognition performance and is highly dependent upon the settings of the other variables.
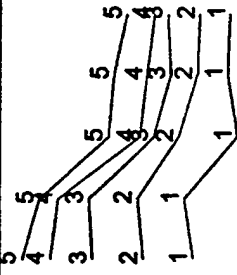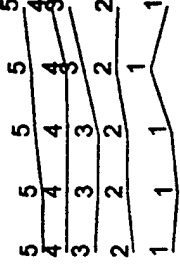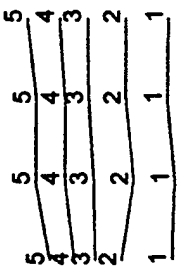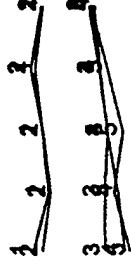
The main effects for each level of filter width increase as the width of the bandpass increases. This is partially due to the spxwd interaction effect mentioned above. The difference between the main effects for the widest and narrowest settings is slightly larger (~9) than the experimental error in the training procedure.

Based on the above discussion, the order for optimizing the filter design variables and a recommended number of variable levels can be determined. The interferogram segment starting position, bandpass position, and bandpass width variables should be studied first. As many levels as possible for interferogram segment starting position should be studied. Because the magnitudes of the main effects for the levels of segment starting position were larger than the main effects of the levels for filter position and width, a greater effort should be used to find the optimal starting position since it greatly affects pattern recognition performance. In addition, interferogram segment starting position is highly dependent on the variable setting for bandpass filter width, thereby justifying the need for additional levels to be studied. The width of the analyte band will determine how close to the centerburst the segment can be located. For a narrow analyte band such as the S-F stretching band at 945 cm^{-1}, an interferogram segment starting at point 100 is close enough to achieve good pattern recognition performance. For wider analyte absorption bands such as the C-O stretching band of methanol, the segment can be located as close as 25 points from the centerburst [11].

Based on the plots in Figure 7, the small main effects associated with the settings for filter position indicate that only a few filter position settings need to be chosen for any optimization study. The bandpass position setting should be chosen to coincide with the targeted spectral band of the analyte, although the optimal position may be shifted from the band center, especially in cases in which nearby spectral bands of interfering species are present [12].

Bandpass filter width is the least predictable filter generation variable because it is so highly correlated with both interferogram segment starting point and bandpass filter position. For example, wider filters may perform better, on average, but the optimal filter may have a narrow bandpass. In this study, the filter associated with the largest response score had a width of 45.4 cm^{-1} (second to narrowest). Figure 7 suggests that on average, however, the widest filter is the best. Thus, in any optimization study, several levels of bandpass filter width should be studied.

Despite one interaction term (slxsp) being significant, we believe segment length can be optimized independently of the other variables. This can be justified by viewing Figure 8. This figure is a matrix of six interaction plots displaying the mean response function scores for all

Figure 8. Matrix of interaction plots displaying the mean response function score for each unique pairing of the four experimental variables. The vertical axis is constant across the plots and is in units of the response function, while the settings on the horizontal axis correspond to each column variable setting. The number assigned to each response curve is the row variable setting.

possible combinations of the two variables involved in each two-way interaction. The vertical scale is the same for all plots in the matrix and is in units of the response function. The horizontal scale displays the levels of the column variable. The lines (response curves) in each plot correspond to the mean response function score at the levels of the row variable. Analogous to Figure 7, the settings for each variable are in numerical order. The plot of a specific interaction is located at the intersection of the row and column of the two variables. For example, the upper leftmost plot is the slxsp interaction plot with the 5 response curves corresponding to the settings for segment length. Each response curve has 5 points corresponding to the different settings for segment starting position.

The interpretation of these plots is straightforward. Parallel response curves within each plot indicate that little or no interaction is present among the variables. It is clear from these plots that the two-way interactions involving interferogram segment length (top row) show response curves in a series of steps with the longer segment length settings having larger mean response function scores. The remaining three plots show significant interaction among the variables. The response curves in these plots are often overlapping and are not parallel. This would indicate that segment length can be optimized independently from the other variables due to the less significant interaction effects.

Further evidence for this conclusion can be seen by studying the slxsp interaction plot in more detail. The response curves for the fourth and fifth segment length settings (i.e., 120 and 140 points) are more parallel than the response curves for the other settings. This indicates that the interaction effects between interferogram segment length and starting position are more significant at the shorter interferogram segment lengths, thereby suggesting that segment length can be optimized independently from segment starting position if segment length is set at 120 points or greater. Since the computed F-values in Table 4 suggest that the interaction effects between segment length and the other two variables are less significant than the interaction with segment starting position, fixing the setting for segment length at 120 points or greater should also effectively minimize the impact of the other interactions. Thus, under this scheme, segment starting point, filter position, and filter width should be optimized together while the setting for segment length is held fixed at 120 points or greater. Segment length would be optimized independently after the optimal settings for the other three variables have been determined.

Because segment length has the most influence on pattern recognition performance, several levels should be included in an optimization study. While shorter interferogram segments are computationally more efficient, longer segments consistently achieve better results. Another important factor to consider in choosing the segment length settings to study is the width of the analyte band. For wider analyte bands, the analyte information is compressed into a smaller region of the interferogram than for narrow analyte bands due to the greater damping rate of the interferogram signal. At a certain segment length, dependent upon the compound being studied, the improvement in pattern recognition performance becomes negligible and further study of longer segment lengths is not needed. Segment lengths of greater than 150 points appear to define the point of diminishing returns for most analytes.

Validation of Protocol. To demonstrate that the conclusions drawn from the SF_6 laboratory data were valid, the filter generation protocol was applied to actual SF_6 field FTIR remote sensing data. This study was performed by calculating four bandpass filters with variable settings that should yield good pattern recognition results and four bandpass filters with variable settings that should yield poor results. The variable settings chosen for the "good"

37

filters were those which achieved the four best pattern recognition results with the laboratory data. One of the interesting observations of this study was that poor pattern recognition results can be obtained by two different strategies. One is to use variable settings corresponding to narrow filters and short interferogram segments located near the centerburst. The other is to use short interferogram segments remote from the centerburst. In either case, the majority of the SF_6 information is missed. Thus, both types of "poor" filters should be represented in the validation of the protocol. Two variable settings were chosen for each type of "poor" filter. The four "good" filters were assigned sequence numbers 1-4, while the "poor" filters were numbered 5-8. These variable settings are shown in Table 5. For each filter, the mean classification percentage for the SF_6 laboratory data is also shown in Table 5. The mean values were computed across the two replicates for each filter. For the laboratory data, the average numbers of patterns correctly classified for the "good" and "poor" filters were 852.5 (96.9%) and 717.3 (81.5%), respectively.

Open Path Interferogram Data Analysis. Field measurements are more challenging than laboratory measurements due to the widely varying background conditions present. The impact of this increased background variation is less separation in the data space between the compound-containing and background interferogram segments. Thus, use of the same training protocol that was employed with the laboratory data will typically result in poorer pattern recognition performances for the field data.

A five-vector discriminant was computed and optimized by use of the same training protocol employed with the laboratory data. The training set of 3000 interferograms was used in positioning the discriminants. Subsequently, the discriminants were applied to the 1000 test interferograms in the prediction set. The training and prediction results are presented in Table 6.

The mean percentages of SF_6-containing patterns correctly classified in training using the open path data were 96.6% for the "good" filters and 85.5% for the "poor" filters. The average percentages of all patterns correctly classified in the prediction set for the open path data were 94.8% for the "good" filters and 89.5% for the "poor" filters. Thus, the "good" variable settings outperformed the "poor" variable settings by greater than 10% in the training step, analogous to the results obtained with laboratory data. These results validate the protocol devised for the filter generation and show that it is applies to open path as well as laboratory data. The occurrence of prediction classification percentages that are greater than the corresponding training percentages is an artifact of this data set and does not affect any of the conclusions drawn.

Conclusions

The results obtained in this study indicate that a protocol for designing a near-optimal bandpass filter can be developed. Conclusions based on the laboratory data and confirmed with the open path data show that interferogram segment length has the greatest impact on pattern recognition performance and can be studied independently from the other variables. The three remaining variables must be studied together due to the large interaction effects present among them. Interferogram segment starting position was shown to be the second most influential variable in terms of pattern recognition performance and should be studied in greater detail than either filter bandpass position or filter width.

Table 5

Filter Variable Settings and Classification Performances for Open Path Data

Filter number	Segment length (sl)	[a]Segment starting point (sp)	Filter position (fp) (cm^{-1})	[b]Filter width (wd) (cm^{-1})	[c]Classification performance (Lab data)
Good					
1	140	100	937.0	45.4	97.4
2	140	75	952.4	74.7	97.0
3	140	100	944.7	74.7	96.5
4	140	75	944.7	74.7	96.5
Poor					
5	60	75	937.0	36.4	77.0
6	60	75	940.9	36.4	78.5
7	60	175	940.9	36.4	85.2
8	60	175	944.7	63.6	85.4

[a] Relative to interferogram centerburst

[b] Full width at half-maximum (fwhm)

[c] Mean classification percentage computed for the two replicates

Table 6

Training and Prediction Classification Performance for Open Path Data

Filter number	Training performance (%)	\rediction performance (%)
Good		
1	95.9	94.6
2	96.8	95.8
3	96.8	94.4
4	96.9	94.3
Poor		
5	82.8	88.7
6	85.0	89.2
7	86.4	89.0
8	87.6	91.0

While the work reported here focused on the detection of SF_6, this protocol should be valid for use in developing a detection scheme for other compounds. Given knowledge about the analyte band position and width, an optimization study of the variables employed in filter generation can be completed in a more efficient manner than was previously done. By studying the variables which have the most impact on pattern recognition performance in greater detail, fewer filters will need to be computed to develop a near-optimal detection scheme.

In a general sense, the conclusions drawn from this study may have an impact on other areas of interferogram-based data analysis. Of particular relevance is the quantitative analysis of FTIR interferogram data. Work in this laboratory has demonstrated the effectiveness of obtaining quantitative information from bandpass filtered interferogram data [29]. The same decisions regarding the design of the optimal bandpass filter and interferogram segment are also pertinent in this work.

Automated Detection of Trichloroethylene by Fourier Transform Infrared Remote Sensing Measurements

Remote sensing measurements based on FTIR spectroscopy are becoming increasingly popular for the remote detection of airborne volatile organic compounds (VOCs) [1]. The target compound for the research described here is trichloroethylene (TCE), a toxic solvent that serves as the focus of significant environmental monitoring efforts [30]. The detection of airborne TCE vapor is thus important in a variety of monitoring applications, and FTIR remote sensing measurements represent one option for implementing an automated TCE detection procedure.

The most flexible application of an FTIR remote sensor is the use of the instrument to view a naturally occurring background source of infrared energy. This "passive" spectral measurement allows the sensor to be a highly portable air monitor that can interrogate the atmosphere over large distances. The principal limitation of this approach, however, is the sensitivity of the measurements to changes in the infrared background emission present in the field of view (FOV) of the spectrometer [1,31-34]. The detected background radiance is the resultant of contributions from the infrared radiation source, the intervening atmosphere, and the instrumental response function of the spectrometer [35]. The spectral features of analyte species are superimposed on this varying background and therefore the background information must be suppressed if the analyte signatures are to be observed reliably. The conventional laboratory approach of collecting a representative background spectrum for use in ratioing out or subtracting the non-analyte features is very difficult given the instability of the background.

The interferogram-based analysis methodology described above is focussed on overcoming this limitation by use of novel data analysis strategies for suppressing the contributions of the infrared background without performing an actual background measurement [4,10,11,35,36]. As described previously, this methodology is based on the direct analysis of short segments of FTIR interferogram data and combines bandpass digital filtering and pattern recognition techniques to achieve the detection of VOCs. The filtering step extracts the analyte signature from collected interferograms while the pattern recognition procedure uses the filtered interferogram data as unique patterns to determine the presence or absence of targeted VOCs. In addition to allowing the extraction of analyte information from a variety of backgrounds, the restriction of the analysis to a short interferogram segment has potential benefits in reducing the data collection requirements of the measurement and in simplifying the design of the instrument.

Previous research in our laboratory has demonstrated the ability of this methodology to extract analyte signatures from interferogram data and thereby eliminate the contributions of terrain backgrounds and adjacent spectral bands of interfering compounds [4,11,36]. In the work

41

presented here, this compound detection problem is made significantly more challenging by combining sky and water backgrounds with the terrain backgrounds used previously. The ability of the methodology to implement an automated detection of TCE in the presence of this extreme background variation is assessed.

Experimentation

Instrumentation. The FTIR spectrometer employed in this work was a Brunswick emission spectrometer. (Brunswick Technical Group, DeLand, FL). The modulator for the spectrometer is based on the flex-pivot 'porch swing' Michelson interferometer design. By use of an interferometer mirror velocity of 1.269 ± 0.017 cm/sec [37], the infrared energy was modulated onto a Hg:Cd:Te detector maintained at 77 K with a Magnavox closed-cycle Stirling cooler. The detector was narrow band and optimized for use in the 8-12 μm atmospheric transmission window. The spectrometer FOV was 1.5° and was reduced to 0.5° with an antireflection coated germanium refractive optic telescope (Intellitec, DeLand, FL) designed for open-air monitoring applications.

The interferometer was interfaced to a Dell System 486P/50 IBM PC compatible computer (Dell Computer, Austin, TX) operating under MS-DOS (Microsoft, Redmond, WA). The data acquisition was performed by use of the MIDAS package [38]. Interferogram points were acquired at every eighth zero crossing of the reference laser, giving a maximum spectral frequency of 1975 cm^{-1}. A total of 1024 sampled interferogram points allowed calculation of spectra with points spaced at approximately 4 cm^{-1}.

Methods. Interferogram data were acquired with three types of passive remote sensing measurements. The three experiments were termed (1) open-air terrestrial, (2) passive cell terrestrial, and (3) passive cell laboratory measurements. The three approaches were used in order to obtain variation in both TCE concentrations and in the infrared backgrounds observed.

In the open-air terrestrial measurements, TCE vapor was released into the atmosphere with an evaporative emission source for which the concentrations were not controlled. This measurement restricted the spectrometer FOV to ensure that the vapor cloud filled the FOV. The spectrometer was located from 5 to 25 m from the vapor emission source. Radiances from terrain, water, low-angle sky, or some combination of both sky and terrain served as the infrared backgrounds for these experiments.

In the passive cell terrestrial experiments, pure TCE vapor was introduced into a gas cell and various terrain, water, and sky backgrounds were viewed through the cell. For these measurements, the cell was held by a bracket attached to the spectrometer housing. The TCE concentrations were not controlled. The cell used had a 62 cm^2 aperture and a 8.2 cm path length. The cell windows were composed of low density polyethylene (0.0005 in. thickness).

The passive cell laboratory experiments employed the same gas cell. In these measurements, the concentrations of TCE were obtained by evaporation of various solution mixtures of TCE in carbon tetrachloride (CCl$_4$) [39]. The solutions were prepared by mixing reagent grade TCE (Aldrich, Milwaukee, WI) and reagent grade CCl$_4$ (MCB Manufacturing, Cincinnati, OH). The volume fraction of TCE in CCl$_4$ ranged from 1 to 1/64, corresponding to vapor pressures of 69.2 and 0.95 Torr, respectively, at 25 °C [40]. The vapor pressures of TCE were computed by use of the Wilson equation, converted to ppm by assuming ideal gas behavior, and scaled by the cell path length to obtain path averaged concentrations in units of ppm-m. For TCE volume fractions of 1, 1/2, 1/4, 1/8, 1/16, 1/32, and 1/64, the corresponding path-averaged concentrations were 7466, 3679, 1748, 842, 410, 205, and 102 ppm-m,

respectively. A 4x4-inch laboratory extended blackbody infrared source (Model SR-80, CI Systems, Agoura, CA) was viewed through the cell to simulate conditions found in open-air monitoring applications. The temperature of the source was varied from 5 to 50 °C with an accuracy of ±0.03 °C and a precision of ±0.01 °C.

Assembly of Data. In this discussion, interferograms containing TCE signatures will be termed "TCE-active" interferograms, while those containing no evidence of TCE presence will be termed "TCE-inactive" or "background" interferograms. All interferograms collected with TCE in a cell were deemed to be TCE-actives and used in assembling the data sets employed in developing and testing the TCE detection algorithm. Visual inspections of spectra corresponding to the lowest TCE concentrations in the passive cell laboratory data revealed no evidence of the TCE spectral bands at 845 and 938 cm^{-1}. This confirmed that TCE signals at or near the limit of detection were present in the data set. The open-air terrestrial data were examined for the presence of both TCE spectral bands. The collected interferograms were Fourier processed to single-beam spectra, followed by subtraction of a similarly processed single-beam background spectrum known to contain no TCE features. If clear visual evidence of both TCE spectral bands was observed, the corresponding interferogram was judged TCE-active and placed in a pool of analyte-active interferograms for possible inclusion in the final data sets. Spectra judged indeterminate in terms of TCE presence were removed from the data analysis entirely. Background interferograms collected when no TCE was present were inspected for data integrity and then placed in a separate pool for possible inclusion in the final data sets. The total number of interferograms considered for use was 159,002.

The pools of inspected TCE-active and background interferograms were used to assemble two separate training and prediction data sets for use in evaluating the TCE detection methodology. The two training data sets were used to optimize the digital filtering and pattern recognition parameters of the TCE detection algorithm. The prediction data sets were withheld from these optimizations and were employed subsequently as independent test sets for use in determining the rate of positive and false detections afforded by the optimized detection algorithm. A subset selection procedure developed by Carpenter and Small was used for the selection of training and prediction sets that were representative of the total set of interferograms [7]. Both training and prediction data sets had TCE-active and TCE-inactive interferograms.

The first data set assembled was used to evaluate the ability to detect TCE against infrared backgrounds similar to those we have encountered previously in work with other compounds. The details of the training and prediction sets comprising data set A are presented in Table 7. The training data set had 49,152 interferograms, while the corresponding prediction set had 60,000 interferograms. As indicated in Table 7, the data sets contained interferograms collected with blackbody, terrain, and water backgrounds, some of which included manmade objects such as buildings and vehicles in the FOV. In addition, acetone, methyl ethyl ketone (MEK), and sulfur hexafluoride (SF_6) were present during the collection of some of the field background interferograms to serve as potential interferences. SF_6 has an absorption band centered at 945 cm^{-1} which overlaps to a large extent with the 938 cm^{-1} band of TCE. The primary absorption bands of acetone (1217 cm^{-1}) and MEK (1175 cm^{-1}) do not overlap with either of the TCE absorption bands, but they provide a means for testing the spectral selectivity of the interferogram-based analysis. In the passive cell laboratory data, CCl_4 was present with TCE during the collection of all of the analyte-active interferograms except for those with a TCE volume fraction of 1. Some of the background interferograms were collected with pure CCl_4. The CCl_4 band centered in the region of 790 cm^{-1} does not overlap with the TCE band at 845 cm^{-1}, but it does provide a further test of the frequency selectivity of the filtering procedure.

Table 7

Description of Data Set A

Type of interferograms	Training set		Prediction set	
TCE-actives				
Passive cell laboratory (with CCl_4)		6000		4624
Open-air/passive cell terrestrial				
TCE with building/tree	10040		2919	
TCE over water	924		90	
TCE with blackbody	3036		119	
Sub-total		14000		3128
TCE-inactives				
Passive cell laboratory				
No chemicals	316		36366	
Pure CCl_4	184		1605	
Sub-total		500		37971
Open-air/passive cell terrestrial				
No chemicals	3640		8566	
Acetone over water	1085		2361	
Acetone with building/tree	11078		1282	
Acetone with blackbody	4842		861	
MEK with building/tree	7212		1037	
MEK with blackbody	224		135	
SF_6 with vehicle	571		35	
Sub-total		28652		14277
Total		49152		60000

One of the goals of this research was to test the interferogram-based methodology with the most challenging data set possible. Data set B, described in Table 8, was assembled to meet this goal. The training and prediction sets contained 10,000 and 60,000 interferograms, respectively. This data set incorporated low-angle sky backgrounds (elevations 20° and 45° from the horizontal) into the same mix of backgrounds employed in data set A. This is the first time that interferograms with sky backgrounds have been used in testing our compound detection methodology.

Data Analysis. Analysis of data set A was achieved by implementing the pattern recognition code using a distributed computing model. To achieve the best overall performance, the iterative part of the algorithm was executed on a Silicon Graphics Onyx R4400 computer (Silicon Graphics, Mountain View, CA) while a section of the algorithm that could be parallelized was implemented on a Thinking Machines CM-5E system equipped with 32 nodes (Thinking Machines, Bedford, MA). The parallel code was written in CM FORTRAN, the data-parallel FORTRAN language for the CM-5E.

In addition to the CM-5E/Onyx implementation, a similar approach was employed for joint use of the Onyx with a Maspar MP-2 equipped with 16,384 nodes (Maspar, Sunnyvale, CA). The parallel code was ported to MPL, the data-parallel C language on the Maspar. For the MP-2 code, it was important for performance reasons to have a data set size that was a multiple of 16,384. For this reason, the size of the training set for data set A was set at 49,152. The 2560 pattern recognition runs completed were split between the CM-5E/Onyx and MP-2/Onyx implementations. Precision tests indicated that results obtained with the two computer configurations were effectively identical.

Analysis of data set B was performed on a Silicon Graphics 4D/460 R3000 computer using the Irix operating system (version 5.2). A single processor was employed. The data analysis software was written in FORTRAN 77 and compiled with version 4.0.1 of the Silicon Graphics FORTRAN 77 compiler (optimization level 3). Fourier transform and multiple linear regression computations performed as part of the digital filter design work used subroutines from the IMSL library (IMSL Inc., Houston, TX).

Results and Discussion

Spectral Characteristics. In this study, the two spectral bands of TCE centered around 845 and 938 cm^{-1} were used as the basis for detecting the compound. The band in the region of 845 cm^{-1} arises from an in-plane asymmetric rotational twist around carbon centers while the band near 938 cm^{-1} is due to an out-of-plane bending mode of the entire molecule [41,42]. The full width at half height (FWHH) of the bands at 845 and 938 cm^{-1} are approximately 26 and 28 cm^{-1}, respectively.

Figure 9A is an example single-beam spectrum obtained by Fourier processing an interferogram collected with a tree background when TCE was present in the FOV of the spectrometer. This figure shows the characteristic detector response envelope associated with the spectrometer. The dip in the spectrum in the region of 900 cm^{-1} is a feature of the detector. No TCE absorption features are visible in the single-beam spectrum due to their small magnitudes. Figure 9B is a similar single-beam spectrum obtained by use of an interferogram collected when TCE was viewed against a sky background. A large number of spectral features are observed superimposed on the detector envelope due to the presence of trace atmospheric species viewed over a long path length. No TCE features can be observed in the single-beam spectrum. A principal challenge in the work presented here was the inclusion in the same data set of infrared backgrounds that vary as much as those depicted in Figure 9.

Table 8

Description of Data Set B

Type of interferograms	Training set		Prediction set	
TCE-actives				
Passive cell laboratory (with CCl_4)		1000		6000
Open-air/passive cell terrestrial				
TCE with building/tree	484		6809	
TCE over water	77		582	
TCE with blackbody	325		2899	
TCE with sky	1614		3710	
Sub-total		2500		14000
TCE-inactives				
Passive cell laboratory				
No chemicals	595		11849	
Pure CCl_4	405		1151	
Sub-total		1000		13000
Open-air/passive cell terrestrial				
No chemicals	1322		7506	
Acetone over water	552		3489	
Acetone with sky	1621		3191	
Acetone with building/tree	281		5041	
Acetone with blackbody	32		993	
MEK with building/tree	1275		4307	
MEK with sky	381		2078	
SF_6 with vehicle	36		395	
Sub-total		5500		27000
Total		10000		60000
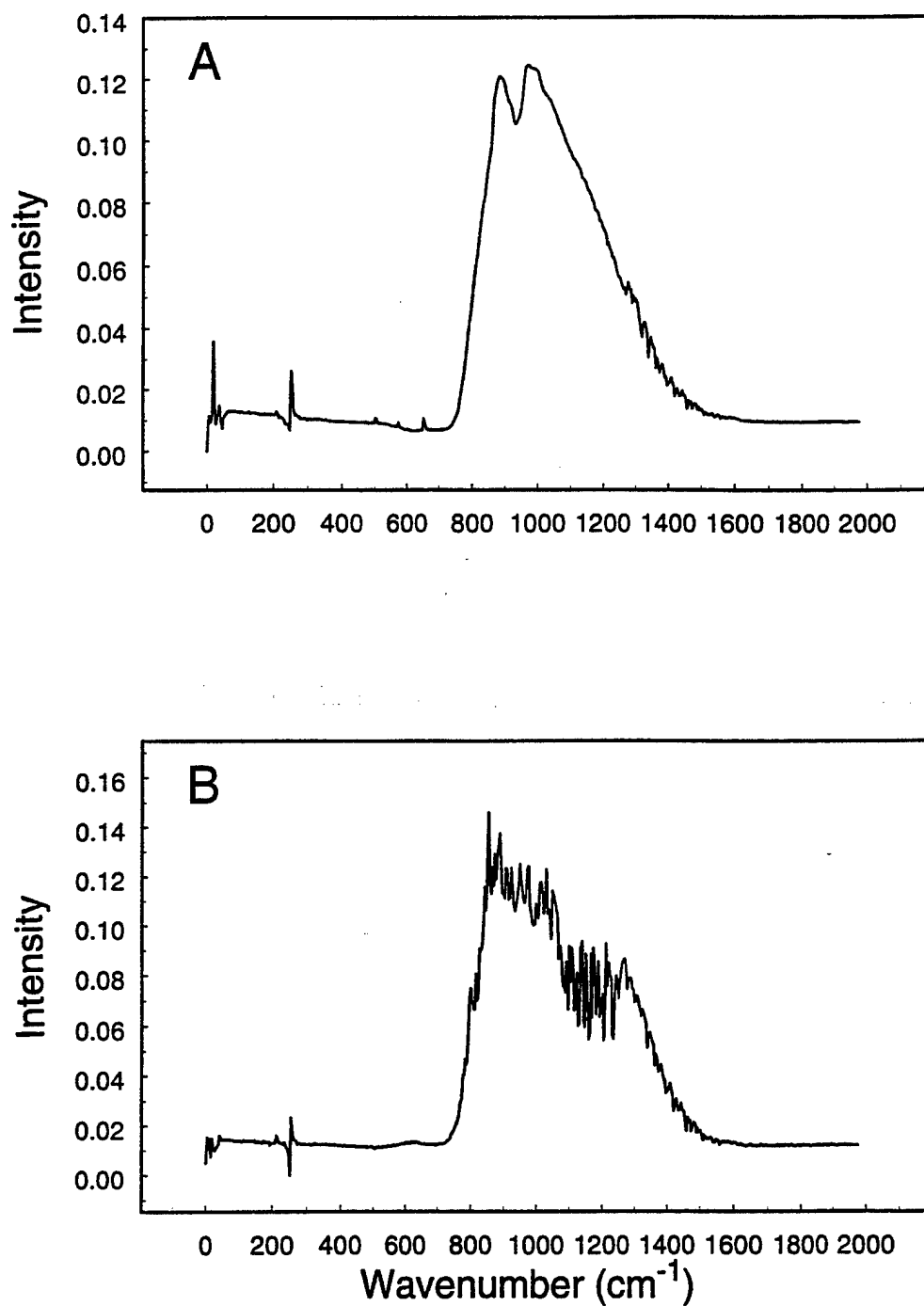
Figure 9. Single-beam FTIR spectra collected with tree (A) and low-angle sky (B) infrared backgrounds in the FOV of the spectrometer. The many narrow features in the sky background derive from the observation of trace atmospheric species over a long path length.

47

Figures 10A and 10B are examples of TCE absorbance spectra. These spectra were obtained by ratioing computed single-beam spectra collected when TCE was present to similarly processed background spectra and converting the resulting transmittance values to absorbance. Although Figure 10 shows examples of absorption features only, emission spectral features are also encountered routinely in remote FTIR measurements. The plotted spectra illustrate the wide range of spectral band intensities encountered. Figure 10A is an example of an intense TCE spectrum, with peak absorbances of approximately 0.17 and 0.16 absorbance units (AU) for the 845 and 938 cm^{-1} bands, respectively. By contrast, Figure 10B is an example of a weak TCE spectrum, with peak intensities of approximately 0.0022 and 0.0020 AU for the corresponding bands. Figures 9 and 10 illustrate that an effective TCE detection method must be able to extract a large dynamic range of TCE absorption and emission signals from infrared backgrounds that exhibit tremendous variation.

Overview of Data Analysis Methodology. The interferogram-based methodology used in this work was developed in our laboratory to extract analyte spectral features directly from FTIR interferogram data. As described previously, the approach is based on the application of a bandpass digital filter to a segment of the interferogram to isolate specific frequencies associated with a spectral band of the target analyte. Pattern recognition methods are applied subsequently to the filtered interferogram segment to implement an automated yes/no decision regarding the presence of the analyte. A separate background or reference measurement is not required because the combination of a judicious choice of interferogram segment and the application of the bandpass filter serve to remove the analyte feature from the background.

The digital filtering is tailored to extract only the frequencies contributing to the spectral bands of the analyte. For TCE, the frequencies contributing to the intensity of spectral bands centered at 845 or 938 cm^{-1} are selectively extracted from the corresponding interferograms by applying a digital filter tailored to the specific spectral region of interest. The resulting filtered interferogram is reduced to a superposition of a series of sinusoidal waves corresponding to the specific frequencies passed by the filter. The filter passband characteristics such as position and width for a Gaussian-shaped filter are specified in the spectral (i.e., wavenumber) domain and the filter is generated for application in the interferogram (i.e., spatial) domain by use of suitable design methods. The interferogram-domain filter approximates the frequency response specified in the wavenumber domain. The digital filter design technique employed in this study is an implementation of a finite impulse response (FIR) digital filter with time-dependent coefficients [8].

Application of the filter rejects background information located at frequencies outside the passband of the filter. In effect, the infrared background emission is truncated to the shape of the passband of the filter. Within the passband, the broad infrared background emission is removed by windowing the interferogram to isolate a segment displaced from the centerburst region. The purpose of windowing can be understood by realizing that the interferogram representation of the broad background emission damps faster than the corresponding representations of narrower spectral features. Past some point in the interferogram, the representation of the broad background has largely damped to zero, while the signatures of narrower features still have significant amplitude. Thus, by making judicious choice of the interferogram segment for use in the data analysis, much of the infrared background information can be eliminated.

Through this procedure, TCE signatures can be isolated directly from an interferogram that is too short for use in obtaining an accurate spectrum via the Fourier transform. Due to the intrinsic assumption of the Fourier transform that the signal is sampled over infinite time, extremely distorted spectra are produced from short interferogram segments of 50-200 points.
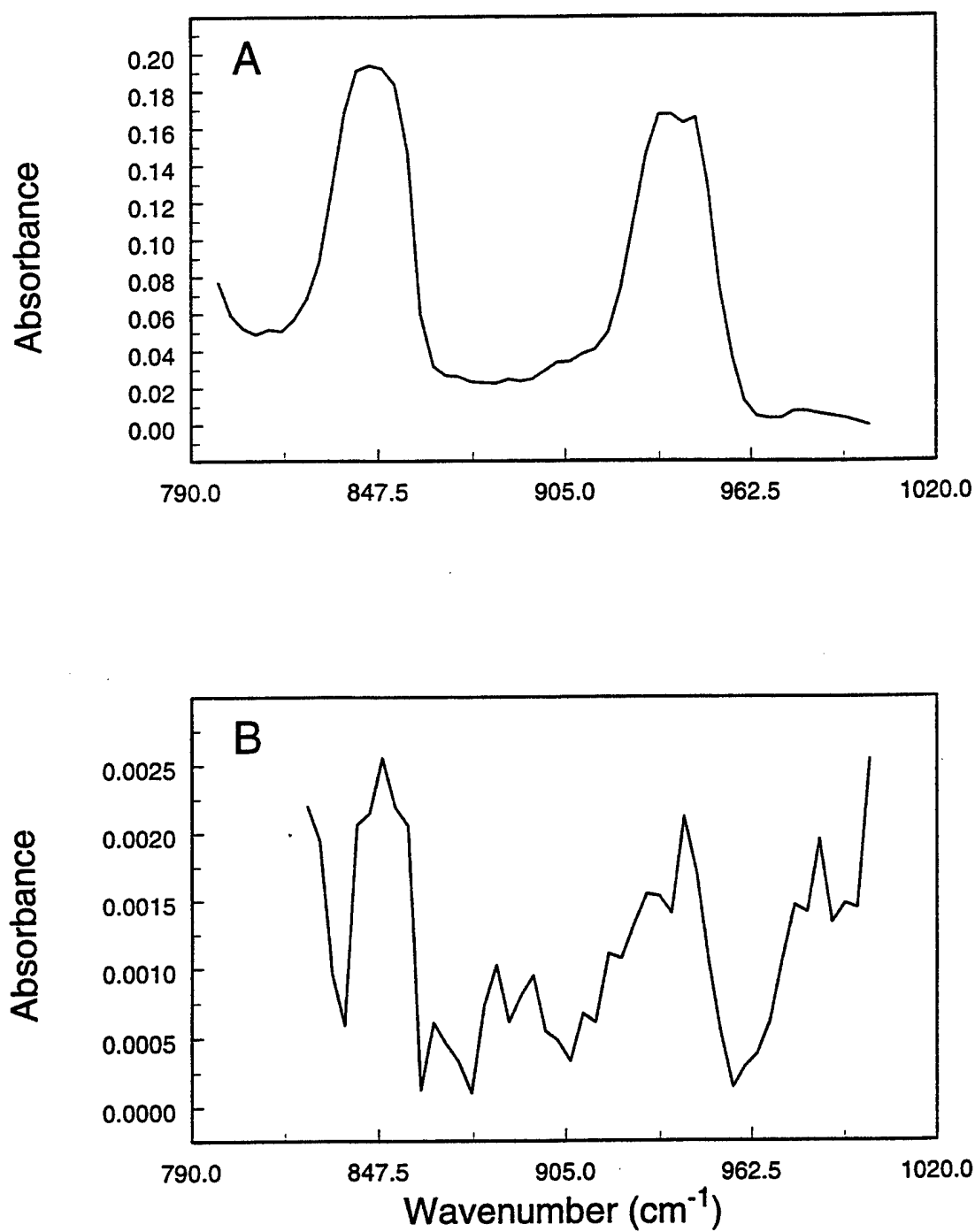
Figure 10. Absorbance spectra of TCE computed from interferograms collected during remote sensing measurements. Both strong (A) and weak (B) TCE spectra are encountered during these measurements.

49

Application of the bandpass filter directly to the interferogram allows frequency selectivity to be incorporated into the analysis without the necessity of meeting the requirements of the Fourier transform.

The next step in the algorithm is to determine whether the information extracted by the digital filter is actually due to the analyte rather than to some interfering species with a spectral band located in the passband of the filter. This is implemented in the pattern recognition step. The position of the targeted spectral band within the digital filter bandpass and the shape of the band are both encoded in the profile of the filtered interferogram segment. The unique profile reflecting both the intensity and shape of the filtered interferogram segment is utilized by the pattern recognition methodology in deciding whether the analyte is present or not.

Filtering an interferogram segment of p points produces a filtered intensity value for each point. The filtered intensities can be considered as a p-dimensional vector characterizing the filtered interferogram segment. This vector can be represented as a point or "pattern" in a p-dimensional space, where the coordinate axes correspond to intensities of specific points in the filtered interferogram. If the filter is effective in extracting the interferogram-based representation of the TCE spectral band, then the data space formed from filtered interferogram segments should contain separate clusters for TCE-active and TCE-inactive segments. The clustering of these segments in the data space according to TCE presence allows the use of pattern classification techniques to make a yes/no decision regarding the presence of TCE information in any filtered interferogram.

The choice of an appropriate pattern recognition method depends on the distribution of points in the p-dimensional data space. Previous work has demonstrated that the data space formed from the filtered interferogram segments is characterized by apparent nonlinear boundaries between the analyte-active and analyte-inactive patterns [9]. Achieving the greatest sensitivity in discriminating the TCE-active and TCE-inactive patterns thus requires a pattern recognition method that can accomodate these nonlinear boundaries. Two such methods are piecewise linear discriminant analysis (PLDA) [9,13] and artificial neural networks [43].

As described earlier, PLDA is based on the construction of separating surfaces or discriminants that define regions of the data space occupied by points belonging to specific categories of data (e.g., TCE-active and TCE-inactive). A numerical optimization procedure is used to find the optimal locations of these separating surfaces. Unknown points corresponding to new filtered interferograms can be classified or assigned membership to one of the data categories by computing the orientation of the point relative to the discriminants. PLDA employs multiple linear hyperplanes to approximate a nonlinear surface separating the different data categories. The number of hyperplanes to employ is the principal configuration parameter associated with PLDA.

Neural networks are a general class of nonlinear modeling techniques that can be adapted to pattern classification applications. The output of the network can be used to assign an input pattern to the data categories being modeled by the network. The network output is generated by transforming the input pattern (e.g., a filtered interferogram segment) through a series of linear and nonlinear functions. The network can be configured in a very general way to incorporate greater or lesser degrees of nonlinearity in generating the output. This flexibility gives the network excellent capabilities in modeling nonlinearities, but it also means that configuring the network can be a challenging optimization problem.

PLDA was selected for use in the TCE detection problem due its simpler configuration requirements. Given the large sizes of data sets A and B and the fact that several variables in

50

the analysis already require optimization (e.g., the interferogram segment and bandpass filter specifications), it was judged undesirable to have to spend additional computational time in seeking the optimal network configuration. Furthermore, since the neural network optimization is typically initialized with a random network, the optimization is very susceptible to the starting conditions. Replication of the optimization with different initial networks is typically performed to overcome this problem. This requirement further increases the computational requirements associated with the use of neural networks in the TCE detection problem. This requirement for replication can be overcome in PLDA by the use of a direct calculation that provides a good starting point for the discriminant optimization [9].

PLDA was implemented in the TCE detection by use of a stepwise procedure to compute the individual discriminants that approximate the nonlinear separating surface between the TCE-active and TCE-inactive data classes. The first discriminant was optimized to classify correctly as many TCE-active patterns as possible. In optimizing the discriminant placement, misclassified TCE-inactive patterns were heavily penalized, resulting in a discriminant that is said to be "single-sided." This discriminant defines a boundary in the data space in which only TCE-active patterns lie on the "pure" side of the boundary and a mix of TCE-active and TCE-inactive patterns may lie on the "mixed" side. When the discriminant position had been optimized, the TCE-active patterns on the pure side of the discriminant were removed from consideration, and a second discriminant was optimized to separate additional TCE-active patterns from the mixed group. This procedure was continued until no more TCE-active patterns could be separated (i.e., until no additional single-sided discriminants could be generated) or until a maximum of four (data set A) or five (data set B) discriminants was reached.

After the discriminants are positioned, any p-dimensional filtered interferogram segment can be classified as either TCE-active or TCE-inactive by the calculation of its discriminant score. The discriminant score is a threshold value computed by the application of the piecewise linear discriminant to the filtered interferogram segment. Filtered interferogram segments with a discriminant score greater than zero are judged to contain the TCE signature, based on the orientation of the vector representation of the segment relative to the previously computed separating surface. The magnitude of the discriminant score indicates the distance of the filtered interferogram segment from the separating surface.

Overview of Parameter Optimization for TCE Detection. The interferogram-based methodology described above was applied to develop an automated detection algorithm for TCE. The variables optimized in this study were the bandpass filter position, bandpass filter width, interferogram segment length, and segment location. The two filter variables determine the frequency range over which the filter operates, thereby defining the degree of spectral selectivity incorporated into the interferogram-based analysis. The interferogram variables determine the selectivity of the analysis with respect to spectral band width, thus helping to remove the effects of broad features of the infrared background. The four variables were optimized for both absorption bands of TCE.

In the work described earlier, it was established that filter position, filter width, and interferogram segment location must be optimized in a joint experimental design due to relationships that exist among these variables. It was also found that as long as a segment length of 120 points or greater was employed in the optimization of the other three variables, the segment length variable could be optimized alone after the optimal values for the other variables had been determined.

Analysis of Data Set A. For data set A, a factorial experimental design study was conducted to optimize the filter position, filter width, and interferogram segment starting point.

Based on the optimization protocol described above, the interferogram segment size was held constant at 121 points during this optimization. A set of 128 filters was generated at sixteen filter positions from 848.5-964.2 cm^{-1} and with eight different nominal filter widths ranging from 85-200 cm^{-1} (FWHH). Twenty segment locations were used. The first segment was 40/160 where the numerator denotes the starting segment point and the denominator indicates the ending point of the segment relative to the interferogram centerburst. This segment window was moved by five points until the last segment of 135/255 was reached, resulting in a total of 20 segments tested. All combinations of the three variables were investigated by performing the PLDA calculation described above. This resulted in a total of 16 x 8 x 20 = 2560 PLDA runs. For each run, the training set was used to attempt to compute a piecewise linear discriminant consisting of four individual discriminants. For 139 of the 2560 runs, the variable settings performed so poorly that the first discriminant of the piecewise linear discriminant could not be made single-sided. These combinations were removed from further analysis, leaving 2421 valid discriminants. For 53 cases, only the first discriminant could be made single-sided, while for 76 cases only two single-sided discriminants could be computed. For 117 additional cases, only three single-sided discriminants were obtained. This left 2175 cases in which all four discriminants were single-sided. In evaluating the results for the 2421 valid discriminants, the maximum number of single-sided discriminants was used.

The 2421 valid discriminants were applied to the 60,000 interferograms in the prediction set. These interferograms were not used at any prior stage during the filter generation or the computation of the discriminants. The prediction results were observed to track the training results almost exactly. This is due to the use of a training set which was large enough to be globally representative of the various terrain, water, and blackbody backgrounds encountered.

The percentage of correctly classified TCE-active interferograms and the percentage of false detections were studied as a function of the FWHH of the filter bandpass and the starting point of the interferogram segment. Many different combinations of filter width and segment location were observed to achieve both a high TCE detection percentage and a low rate of false detections. The only poor combinations corresponded to the narrower filters coupled with the segment locations closest to the centerburst. This can be rationalized by considering that the narrower filters truncate the infrared background emission to a narrow spectral feature which takes relatively long to damp out in the interferogram. Thus, the TCE signature near the centerburst is still obscured by the background. The most promising combinations corresponded to the narrower filters coupled with segment locations relatively far from the centerburst. These filters and segments achieved both very high detection percentages and a low rate of false detections.

To explore this region further, the 2421 prediction results were reduced to the 212 cases in which the filter FWHH was less than 120 cm^{-1} and the segment starting point was greater than point 100 (relative to the centerburst). The prediction results for these cases were studied as a function of filter bandpass position. This study revealed that TCE information can be extracted reliably with filters positioned across the range of 850-960 cm^{-1}. Furthermore, no trends were noted in the false detection rate with respect to filter position. This suggests that spectral interferences are not contributing to the classification results. In terms of reliable TCE detections, the most stable results were observed with filters positioned from 860-910 cm^{-1}. Given that the filters employed here were on average approximately 100 cm^{-1} wide at the midpoint of the passband, the filters positioned in the 860-910 cm^{-1} region clearly isolate information from both TCE spectral bands. The fact that both spectral bands are of similar width (see Figure 10A) suggests that the TCE information from each band will be represented in the interferogram similarly with respect to location.

Overall, the prediction results obtained with data set A indicate that TCE can be detected with a reliability approaching 99% and with a false detection rate less than 0.5%. The reliability of these detections ranks with the best we have obtained in our remote sensing work. We turn now to the question of whether similarly reliable detections can be made when complex sky backgrounds are incorporated into the data sets.

Analysis of Data Set B. The analysis of data set B consisted of two phases. In the first phase of the study, 125-point interferogram segments were employed. The segment locations studied were 51/175, 76/200, 101/325, and 126/250, where the numerator and denominator indicate the starting and ending points of the segments, respectively, relative to the centerburst. Both 845 and 938 cm⁻¹ TCE spectral bands were used independently for the analysis. Five filter positions and three nominal filter widths were employed for each band. Table 9 lists the filter positions along with the actual FWHH values for the passbands of the computed filters. The FWHH values were computed based on the actual frequency responses of the filters. Due to the presence of sky backgrounds in data set B, no filters centered between the two TCE bands were employed in order to reduce the degree to which spectral features associated with atmospheric species were passed by the filters.

The experimental design described above resulted in a total of 120 piecewise discriminant calculations (4 segments x 5 filter positions x 3 filter widths = 60 experiments for each of the two TCE spectral bands). The computed discriminants were tested through the use of the separate prediction set, and the overall training and prediction results were tabulated.

For the filters based on the 845 cm⁻¹ band, the average training classification result was approximately 96% in terms of the degree of recognition of TCE-active interferograms. The average prediction result was approximately 92%, also in terms of TCE recognition. For the filters based on the TCE band at 938 cm⁻¹, the corresponding average training and prediction results were approximately 96% and 94%, respectively. The rate of false detections in prediction was generally less than 1%.

In the second phase of the study, the interferogram segment location was optimized further in conjunction with the segment length. In this study, the ten best performing combinations of filter position and width found during the initial work were used. These filters are indicated in Table 9. Segment sizes of 50, 70, 90, and 110 points were investigated. Without going past point 240 (relative to the centerburst), the segment starting points were varied from points 51-191 in steps of 20 points. This produced 8, 7, 6, and 5 segments of lengths 50, 70, 90, and 110, respectively. A total of 260 piecewise discriminant calculations were performed for each spectral band of TCE (10 filters x (8 + 7 + 6 + 5) segments). For each spectral band and segment length, Table 10 lists the results for the best performing filter/segment combinations. The best results were identified on the basis of the percentage of overall correct classifications. This is defined as the combined percentage of TCE-actives and TCE-inactives correctly classified. Table 10 also lists the percentages of TCE-actives correctly classified in both training and prediction and the false detection rate. For both bands of TCE, the percentage of TCE-actives correctly classified in training is slightly greater than the corresponding prediction result. The false detection rate is less than 1%. For both TCE bands, the training and prediction results track each other.

From the results in Table 10, the filter/segment combination indentified by a filter position of 939.5 cm⁻¹, filter FWHH of 123.4 cm⁻¹, and segment based on points 111/220 was selected for further study. To provide a more complete picture of the prediction results for this case, the classifications were subdivided on the basis of various chemical species present or the type of infrared background. Table 11 lists the number of prediction set interferograms correctly and

Table 9

Digital Filters Used in the Analysis of Data Set B

Filter number	845 cm^{-1} TCE spectral band		938 cm^{-1} TCE spectral band	
	Filter position (cm^{-1})	FWHH (cm^{-1})	Filter position (cm^{-1})	FWHH (cm^{-1})
1	843.1	86.9	931.8[a]	100.3
2	843.1[a]	82.6	931.8[a]	127.3
3	843.1[a]	113.0	931.8[a]	146.6
4	845.1	69.6	935.8[a]	100.3
5	845.1[a]	69.6	935.8[a]	127.3
6	845.1[a]	95.6	935.8[a]	142.7
7	847.0[a]	95.6	939.5	100.3
8	847.0	69.6	939.5[a]	123.4
9	847.0[a]	75.6	939.5	142.7
10	848.9[a]	78.3	943.4	96.4
11	848.9[a]	78.3	943.4	123.4
12	848.9	82.6	943.4[a]	142.7
13	850.8	104.0	947.3[a]	96.4
14	850.8[a]	69.7	947.3	119.6
15	850.8[a]	86.9	947.3[a]	142.7

[a] Filter selected for optimization of interferogram segment position and length.

Table 10

Pattern Recognition Classification Results for Optimal Filters/Segments

Filter position and FWHH (cm⁻¹)	Interferogram segment location (size)	Training (%)[a]	Prediction (%)		
		TCE-actives	TCE-actives	False detection	Total
845 cm⁻¹ band:					
843.1 82.6	131/180 (50 points)	96.4	92.5	0.7	97.0
850.8 86.9	131/200 (70 points)	96.4	93.9	0.8	97.4
848.9 78.3	131/220 (90 points)	98.0	95.2	0.5	98.1
850.8 69.7	71/180 (110 points)	96.6	94.7	0.7	97.8
938 cm⁻¹ band:					
931.8 127.3	71/120 (50 points)	95.8	92.5	0.5	97.2
947.3 142.7	151/220 (70 points)	97.2	93.9	0.8	97.4
939.5 123.4	131/220 (90 points)	98.1	95.8	0.8	98.1
939.5 123.4	111/220 (110 points)	98.5	96.2	0.5	98.4

[a]Due to the single-sided requirement of piecewise linear discriminants, no false detections occur in training.

55

incorrectly classified for various interferogram categories. The category of interferograms primarily responsible for the missed TCE-actives was the passive cell laboratory data. Only 5389 of the 6000 laboratory interferograms were correctly classified (89.8%). By comparison, 13850 of the 14000 open-air/passive cell terrestrial TCE-active interferograms were successfully detected (98.9%). Of the 611 missed detections among the laboratory interferograms, 308 and 261 interferograms corresponded to path averaged TCE concentrations of 102 and 205 ppm-m, respectively. These were the two lowest concentration levels collected. For the remaining 42 missed detections, 20, 19, 1, and 2 interferograms corresponded to path averaged concentrations of 410, 842, 1748, and 7466 ppm-m, respectively.

To understand these results, the origin of the spectral signal in a passive FTIR remote sensing experiment must be considered. As described by Kroutil, et al. [35], the signal detected by a passive remote sensor at a given wavenumber can be approximated as

$$P = [T_a T_t N_b + (1 - T_a T_t) N_t] B \qquad (6)$$

where P is the power of the light incident on the sensor, T_a is the transmittance of the intervening atmosphere between the infrared background and the sensor, T_t is the transmittance of the target analyte cloud, N_b is the spectral radiance of the background, N_t is the radiance of a perfect blackbody emitter at the same temperature as the analyte cloud, and B is a parameter related to the optical collection efficiency of the sensor. T_t in eqn 6 can be expressed as e^{-acl}, where a is the absorptivity of the analyte, c is the analyte concentration, and l is the optical path length of the analyte cloud. Thus, the analyte-specific information is embodied in T_t, with the strength of the analyte signal determined by the product of a, c, and l. The $T_a T_t N_b$ term describes the absorption of background photons by the analyte, while the $(1 - T_a T_t) N_t$ term describes the emission of photons from the analyte. An inspection of eq 1 reveals that the net analyte signal is keyed to the difference between N_b and N_t. For example, if $N_t = N_b$, eqn 6 reduces to $P = N_t B$. This is the case in which the rates of absorption and emission are identical, thus resulting in no detectable analyte signal. This is the inherent limitation of the passive remote sensing measurement, as the ability to detect an analyte is dependent on the existence of a sufficient difference in radiance between the analyte cloud and the infrared background. As defined by the Planck function, the radiance is determined by the temperature of the emitting blackbody. Thus, in practice, detection is keyed by a difference in temperature between the analyte and the background.

The laboratory interferograms were collected by varying both the concentration and temperature of the background. The analyte temperature was not controlled, but was measured during the data collection. Replicate interferograms were collected at each combination of these variable settings, and these replicates are represented in the prediction set. Thus, for each combination of concentration and temperature difference, the prediction results can be used to compute a classification percentage. For the 102 (circles) and 205 (triangles) ppm-m concentrations, Figure 11 plots the percentage of successful TCE detections vs. temperature difference in °C. Temperature differences less than zero correspond to TCE emission signals, while positive temperature differences indicate TCE absorptions. The effects of temperature difference and concentration on the ability of the interferogram-based algorithm to detect TCE are clearly indicated by the curves in the figure. The horizontal dashed line in Figure 11 marks 95% correct classification. If this is used as a criterion for a successful TCE detection, the curves in Figure 11 indicate that at 102 and 205 ppm-m, respectively, temperature differences of approximately 10.5 and 6.9 °C are required to detect TCE absorptions. These limiting temperatures are indicated in the figure by the vertical dashed lines. Since both larger concentration and a greater temperature difference contribute to an increase in the analyte signal, it is reasonable that a smaller temperature difference is required to detect TCE successfully at the higher concentration.

Table 11

Pattern Recognition Prediction Results by Interferogram Type

Interferogram type and total number	Interferogram segment location (size in points)							
	71/120 (50) Correct	Missed	151/220 (70) Correct	Missed	131/220 (90) Correct	Missed	111/220 (110) Correct	Missed
TCE-actives								
Passive cell laboratory (with CCl_4) 6000	4783	1217	4968	1032	5291	709	5389	611
Open-air/passive cell terrestrial 14000	13715	285	13808	192	13874	126	13850	150
TCE-inactives								
Passive cell laboratory 13000	12959	41	12847	153	12882	118	12960	40
Open-air/passive cell terrestrial No chemicals 7506	7467	39	7474	32	7486	20	7478	28
Acetone over water 3489	3489	0	3477	12	3471	18	3486	3
Acetone with sky 3191	3177	14	3168	23	3172	19	3173	18
Acetone with tree 5041	5027	14	4968	73	4954	87	4991	50
Acetone with BB[a] 993	993	0	993	0	992	1	991	2
MEK with sky 3083	3071	12	3073	10	3076	7	3077	6
MEK with building/tree 3302	3294	8	3293	9	3297	5	3300	2
SF_6 with vehicle 395	332	63	382	13	354	41	361	34

[a] Blackbody infrared source in spectrometer FOV.
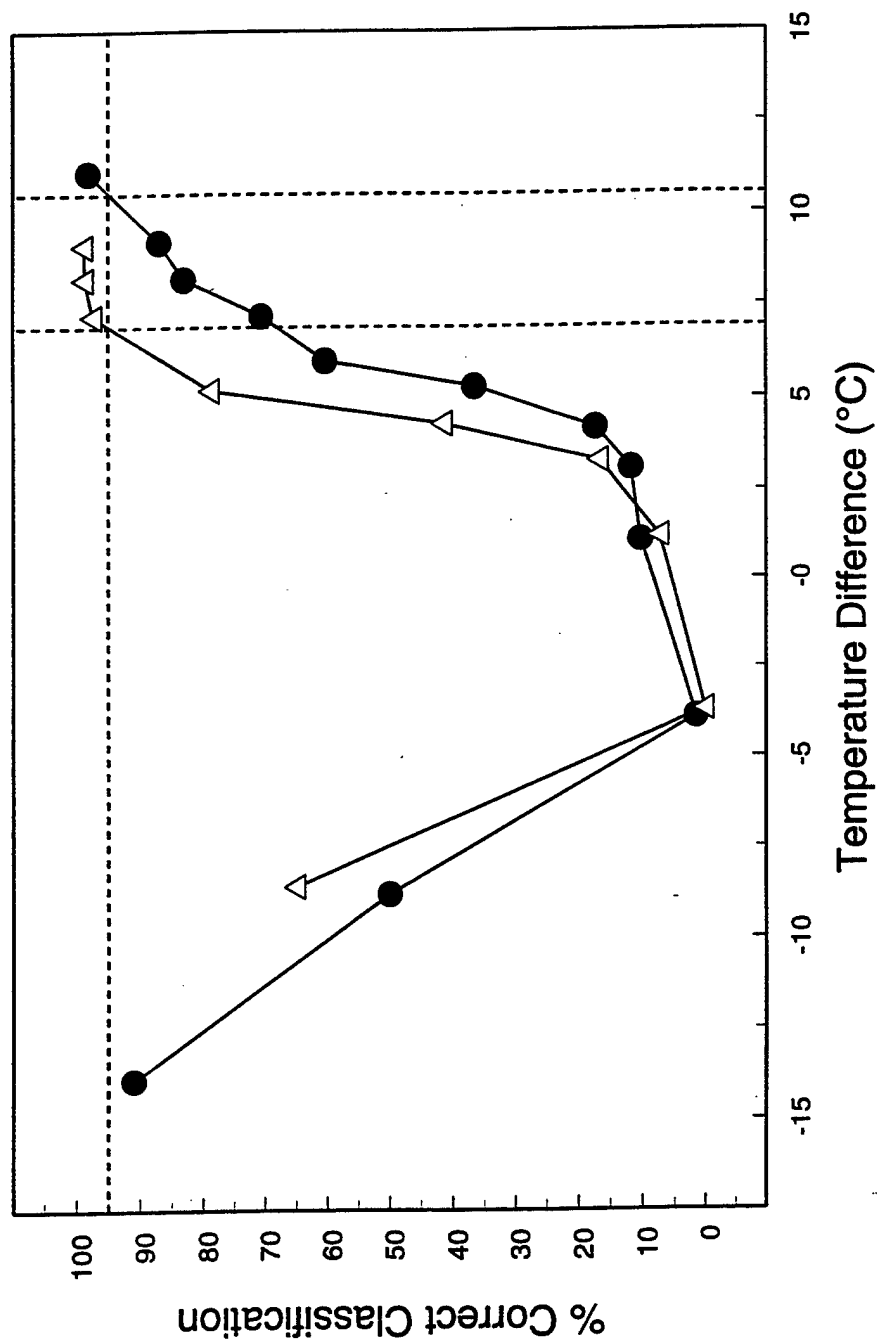
Figure 11. Results from classification of the prediction set of data set B. The filter/segment combination used was filter position 939.5 cm⁻¹, filter FWHH 123.4 cm⁻¹, and segment 111/220. For the path averaged TCE concentrations of 102 (circles) and 205 (triangles) ppm-m, the percentage of the passive cell laboratory TCE-active interferograms correctly classified is plotted vs. the temperature difference (°C) between the infrared background and the analyte. Assuming that 95% correct classification (horizontal dashed line) denotes a successful detection, temperature differences of approximately 6.9 and 10.5 °C (vertical dashed lines) are required for the detection of TCE at 205 and 102 ppm-m, respectively.

As noted above, 42 of the 611 misclassifications among the laboratory interferograms corresponded to the higher TCE concentrations. Here again, the key was the temperature difference. For example, with the 7466 and 1748 ppm-m data, respectively, 93.1% and 98.1% correct classifications were achieved with temperature differences of less than 1.0 °C. For the 410 ppm-m data, however, only 73.9% correct classification was obtained with a temperature difference of 1.3 °C. Thus, defining an effective limit of detection in a passive remote sensing measurement requires consideration of both the concentration and the temperature difference.

Since the laboratory interferograms were collected under conditions of a controlled infrared background, it is possible to compute conventional TCE absorbance spectra by use of background single-beam spectra collected under the same background conditions (i.e., the same N_b) as the TCE spectra. This allowed the computation of spectral signal-to-noise (S/N) ratios for the two combinations of concentration and temperature difference in Figure 11 that are closest to the 95% classification threshold (i.e., 102 ppm-m/11.1 °C and 205 ppm-m/7.2 °C). For each combination, nine interferograms were selected at random from among the replicates in the prediction set and absorbance spectra were computed. For the 938 cm⁻¹ spectral band, a baseline region was defined and second-order polynomial baseline models were computed by polynomial regression. The spectral noise level was computed as the standard deviation of the baseline points about the calculated baseline model. The baseline contribution was subtracted from the spectral band, and the resulting peak maximum was taken as the spectral signal. These signal and noise values were then ratioed to obtain the spectral S/N ratio. For the 102 ppm-m/11.1 °C and 205 ppm-m/7.2 °C combinations respectively, the average S/N ratios of the 938 cm⁻¹ TCE band across the nine spectra were 2.4 and 3.8, respectively. This confirms that the 95% detection threshold in the interferogram-based analysis is occurring in the region of the conventional limit of detection in a spectral analysis (i.e., a S/N ratio of 3.0).

An inspection of Table 5 also reveals that inclusion of the sky backgrounds has not caused an increase in the rate of false detections. The classification percentages for the interferograms collected with sky backgrounds are not significantly different from the results obtained with other background types. The effect of the inclusion of sky backgrounds appears to be a slight reduction in the ability to detect TCE, as evidenced by the lower training and prediction classification percentages for the TCE-active interferograms obtained with data set B relative to the results obtained with data set A. This result is due to the setup of the piecewise linear discriminant calculation to be biased against false detections. As noted previously, the single-sided requirement of the computed discriminants dictates that false detections are heavily penalized in the training set. Thus, increasing the variation among the TCE-inactive interferograms through the inclusion of sky backgrounds results in weak TCE-active interferograms being obscured in the data space. In this case, a discriminant cannot be positioned to separate the weak TCE-actives without causing false detections.

The potential problem of spectral interferences can be studied by considering the false detection rate for interferograms collected in the presence of chemical species other than TCE. Among the species present, the interference due to SF_6 is most pronounced. For the eight filter/segment combinations detailed in Table 10, the SF_6 false detection rate ranges from 2.3-15.9%. Overall, the effect of SF_6 presence is most severe for the filters based on the TCE band at 938 cm⁻¹. This is understandable, given the SF_6 band location of 945 cm⁻¹. As expected, acetone and methyl ethyl ketone do not interfere significantly (false detection rates < 1%) due to the location of their principal spectral bands at 1217 and 1175 cm⁻¹, respectively. The information in these bands has been removed from the interferograms through the use of the bandpass filters.

As noted previously, the implementation of the piecewise linear discriminant procedure produces a discriminant score which indicates the distance of the interferogram segment to the closest discriminant boundary. Positive discriminant scores confirm the segment lies on the TCE-active side of the boundary and negative scores indicate TCE-inactive interferograms. Figures 12-13 illustrate the discriminant score output derived from this procedure for two subsets of the prediction set. Figure 12 plots discriminant scores for 750 open-air/passive cell terrestrial TCE-active interferograms collected with terrain, sky, and water backgrounds, along with 1123 passive cell laboratory TCE-actives collected at different background temperatures and concentrations. Figure 13 is a corresponding plot for 1548 TCE-inactive interferograms with various terrestrial backgrounds. These TCE-inactives included the presence of MEK, acetone, and SF_6 collected with sky and terrain backgrounds.

The various interferogram types are grouped together and labeled in the two figures. The laboratory interferograms in Figure 12 are grouped by concentration and the temperature of the background. During the data collection, the temperature of the background was varied in steps of 5 °C from a minimum to a maximum. For example, the label, 45/10, in Figure 12 indicates that at a path averaged TCE concentration of 7466 ppm-m, the temperature of the blackbody source was varied from 45-10 °C in steps of 5°C.

The results in Figure 12 demonstrate that the discriminant scores for the TCE-active interferograms vary over the range from 0.0 to 0.2. Missed detections are restricted to the region just below the 0.0 threshold. The discriminant scores of the laboratory interferograms clearly contain information about the combination of concentration and temperature difference. Across the collected laboratory interferograms, the temperature of the analyte was relatively constant, ranging from 18-21 °C. Thus, the background temperatures indicated in Figure 12 correlate with absolute temperature difference on either side of the analyte temperature. The discriminant scores approach the 0.0 threshold as the background temperatures approach the analyte temperature. The discriminant scores then increase at background temperatures of 15 and 10 °C as the temperature difference increases and the spectral transitions change from absorption to emission. These trends in the discriminant scores provide further confirmation that the discriminant boundaries lie at the instrumental detection limit of TCE.

Probability-Based Classifications. One disadvantage of the PLDA procedure described above is the arbitrariness of the use of a fixed 0.0 discriminant score threshold in performing a classification. Statistically-based pattern recognition methods have the ability to associate a probability with a classification result on the basis of the distribution characteristics of the training set. However, even though PLDA is a nonparametric pattern recognition method in terms of its empirical placement of the discriminant boundaries, the use of a large data set allows the construction of a reference distribution that can be employed to assign a classificiation probability to the discriminant score of an unknown pattern.

The implementation of a probability-based PLDA classification can be demonstrated by considering that the discriminant scores for the TCE-inactive interferograms in Figure 13 lie in the range of -0.007 to 0.004. The problem discussed previously of false detections in the presence of SF_6 is clearly apparent. However, the majority of the discriminant scores corresponding to the false detections lie very near the 0.0 threshold. This suggests that instead of using 0.0 as the decision threshold, false detections may be reduced by calculation of a probability-based threshold. In such an approach, the decision threshold would be chosen based on the likelihood of a given discriminant score to produce a correct classification.

To test this idea, the 60000 interferograms in the prediction set of data set B were used to construct a reference distribution of discriminant scores. Discriminant scores were obtained
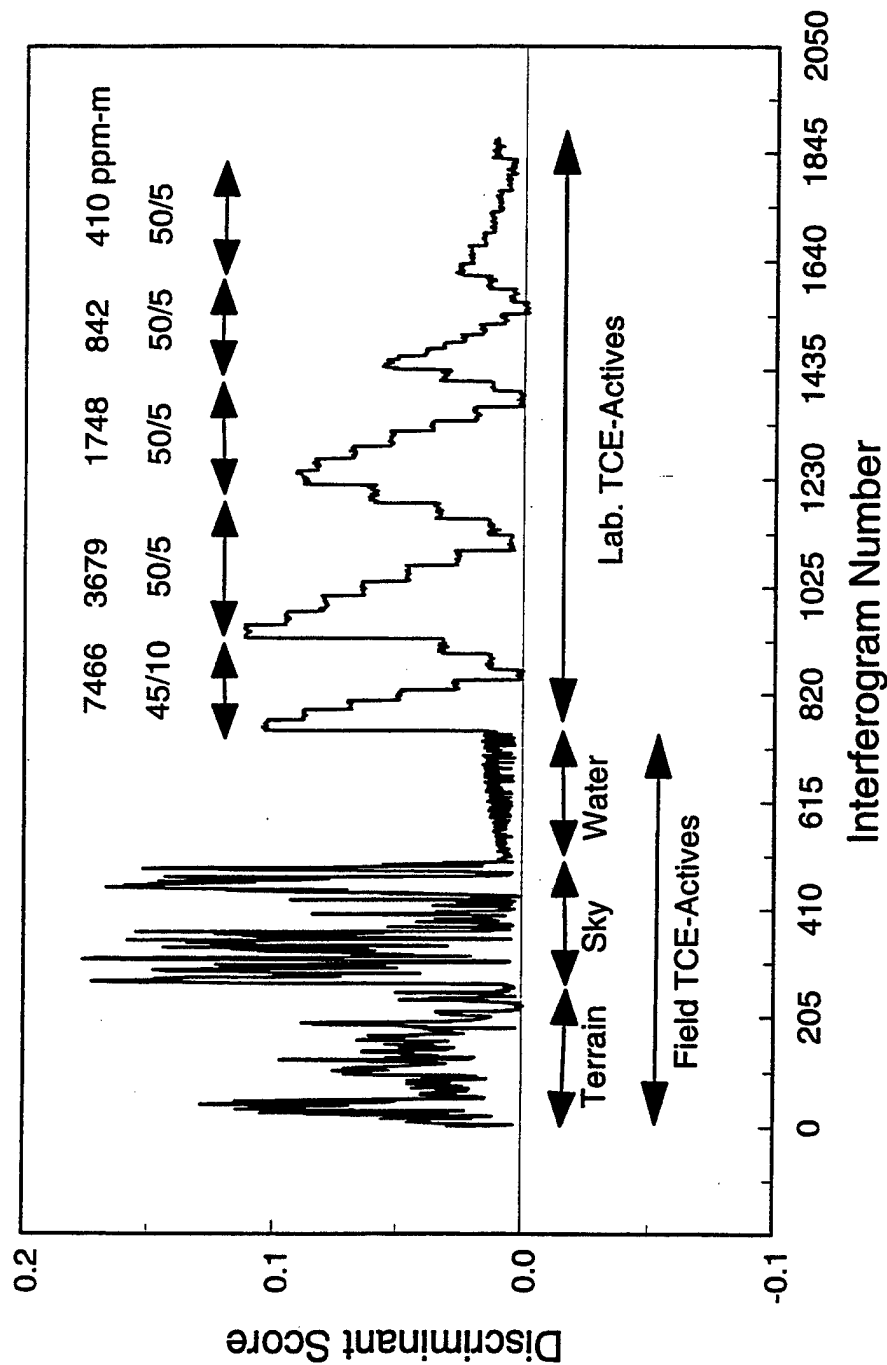
Figure 12. Results from classification of the prediction set of data set B. The filter/segment combination used was filter position 939.5 cm[-1], filter FWHH 123.4 cm[-1], and segment 111/220. Discriminant scores are plotted for a subset of 750 open-air/passive cell terrestrial and 1123 passive cell laboratory TCE-active interferograms. Discriminant scores greater than 0.0 correspond to successful TCE detections. The interferogram types are indicated for the terrestrial backgrounds, and the path averaged concentrations and background temperature ranges are noted for the laboratory interferograms. The temperature ranges denote the starting and ending background temperatures. Steps of 5 °C were used.
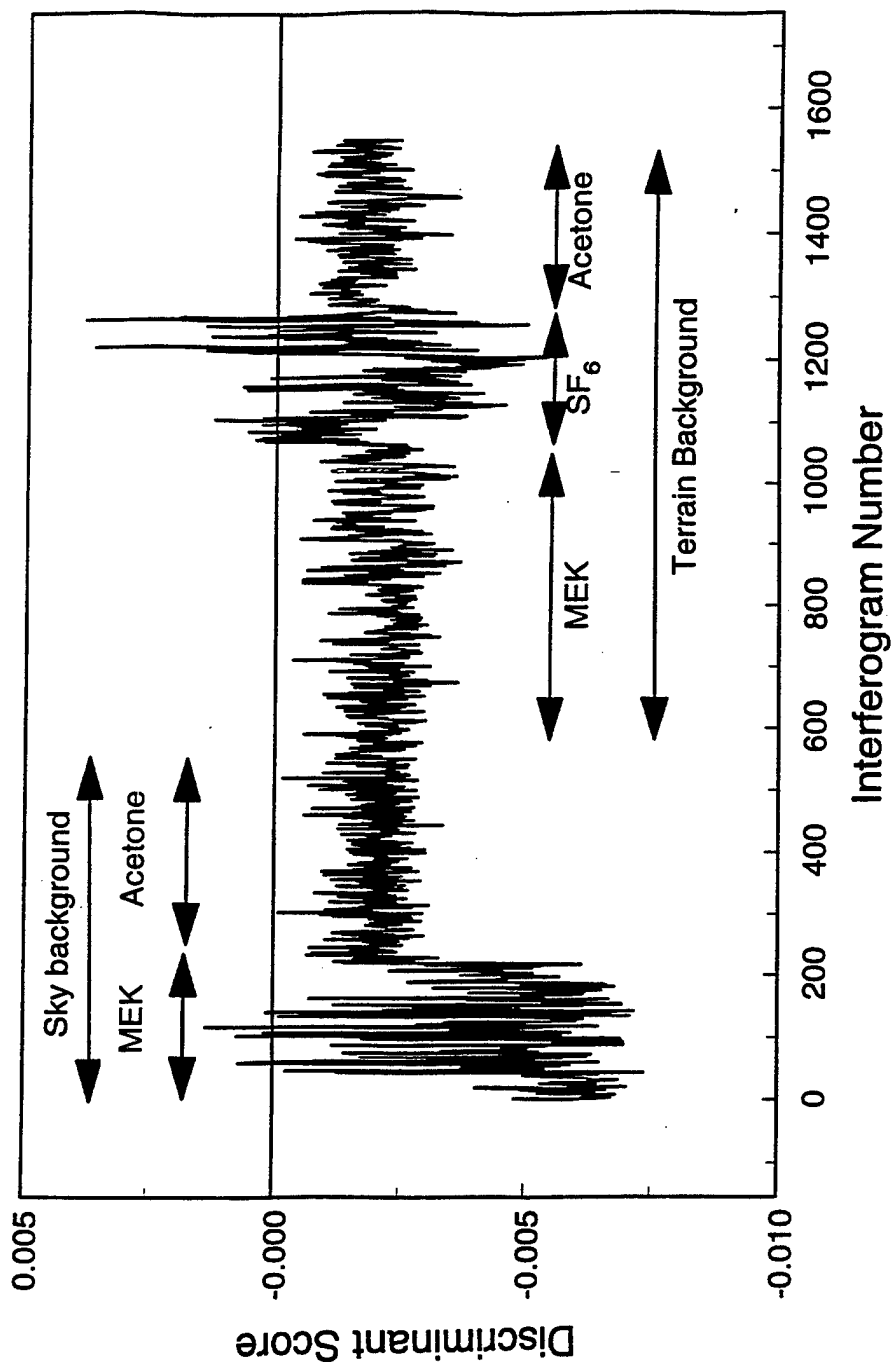
Figure 13. Results from classification of the prediction set of data set B. The filter/segment combination used was filter position 939.5 cm⁻¹, filter FWHH 123.4 cm⁻¹, and segment 111/220. Discriminant scores are plotted for a subset of 1548 open-air/passive cell terrestrial TCE-inactive interferograms. Discriminant scores less than 0.0 correspond to correct classifications, while positive discriminant scores are false TCE detections. The interferogram types are labeled according to the infrared backgrounds observed and the presence of specific chemical species.

by use of the piecewise linear discriminant based on the optimal filter/segment combination described above (filter position 939.5 cm^{-1}, filter FWHH 123.4 cm^{-1}, segment 111/220). These 60000 discriminant scores ranged from -0.0172 to 0.282. This range was divided into bins of size 0.001, and the classification percentages were computed for a given bin based on the correct and incorrect classifications of the interferograms whose discriminant scores fell within that bin. The population sizes of the 301 bins ranged from 0 to 16644. The 21 bins with no interferograms all corresponded to discriminant scores > 0.249. These bins were assigned a classification percentage of 100.0. Figure 14 plots the computed classification percentages vs. discriminant score. As expected, the resulting smooth curve indicates that the percentage of correct classifications decreases near the 0.0 threshold. If 95% correct classification is used as a probability-based threshold, it is observed that the computed classification percentages intersect the threshold at two locations: -0.0017 and 0.0019. In implementing the interferogram-based TCE detection, this discriminant score range could be used as a region of uncertain classification. In effect, this procedure defines a nonparametric statistic that allows a confidence level to be assigned to each classification.

As indicated in Table 11, for the optimal filter/segment combination referenced above, 183 false TCE detections occurred, including 34 false detections when SF$_6$ was present. If the probability-based threshold is used, 141 of the 183 false detections are eliminated. For the SF$_6$ case, 28 of the 34 false detections are avoided. The overall false detection rate is reduced to 0.1% (42/40000), and the rate of false detections due to SF$_6$ is reduced from 8.6 to 1.5%. The use of this criterion does reduce the sensitivity of the algorithm to TCE, however. An additional 1657 TCE-active interferograms would be classified as uncertain, based on the 95% classification probability. Overall, however, the results presented in Figure 14 do confirm that it is possible to assign probabilities to the TCE detections on the basis of discriminant scores. Depending on the needs of the specific monitoring application, these probabilities can be used to strike a balance between TCE detection sensitivity and the false detection rate.

Analysis of Spectral Data. To provide a comparison to the results of the interferogram-based analysis, pattern recognition was also performed on the single-beam spectral data corresponding to the interferograms in data set B. The single-beam spectra were computed by Fourier processing the corresponding interferograms. Triangular apodization and Mertz phase correction were employed. Three different spectral ranges were used, corresponding to 800-1000, 800-1200, and 800-1350 cm^{-1}. Given the nominal 4 cm^{-1} spectral point spacing, these ranges corresponded to 53, 105, and 144 spectral data points. The ranges selected represent different subsets of the detector response envelope depicted in Figure 9. Both TCE spectral bands are included in each range.

The same PLDA training procedure used with the interferogram data of data set B was employed with the spectral data. Piecewise linear discriminants based on five individual linear discriminants were computed. The computed discriminants were then applied to the prediction set. The first three rows of Table 12 list the training and prediction results obtained with the spectral data. Improved performance is noted as the spectral range widens, but in each case, the overall ability to recognize the TCE signature is poorer than with the interferogram-based method.

The principal limitation of using the single-beam spectra for pattern recognition is that the TCE signature represents a very small component of the overall spectrum. As illustrated in Figure 9, no visible TCE signals can be observed in the single-beam spectra. However, just as filtering techniques can be used to extract analyte information from the interferogram, filters can be applied to the spectral data to discriminate against unwanted signals. For completeness, two filtering strategies were applied to the spectral data. First and second-derivative filters based on

63
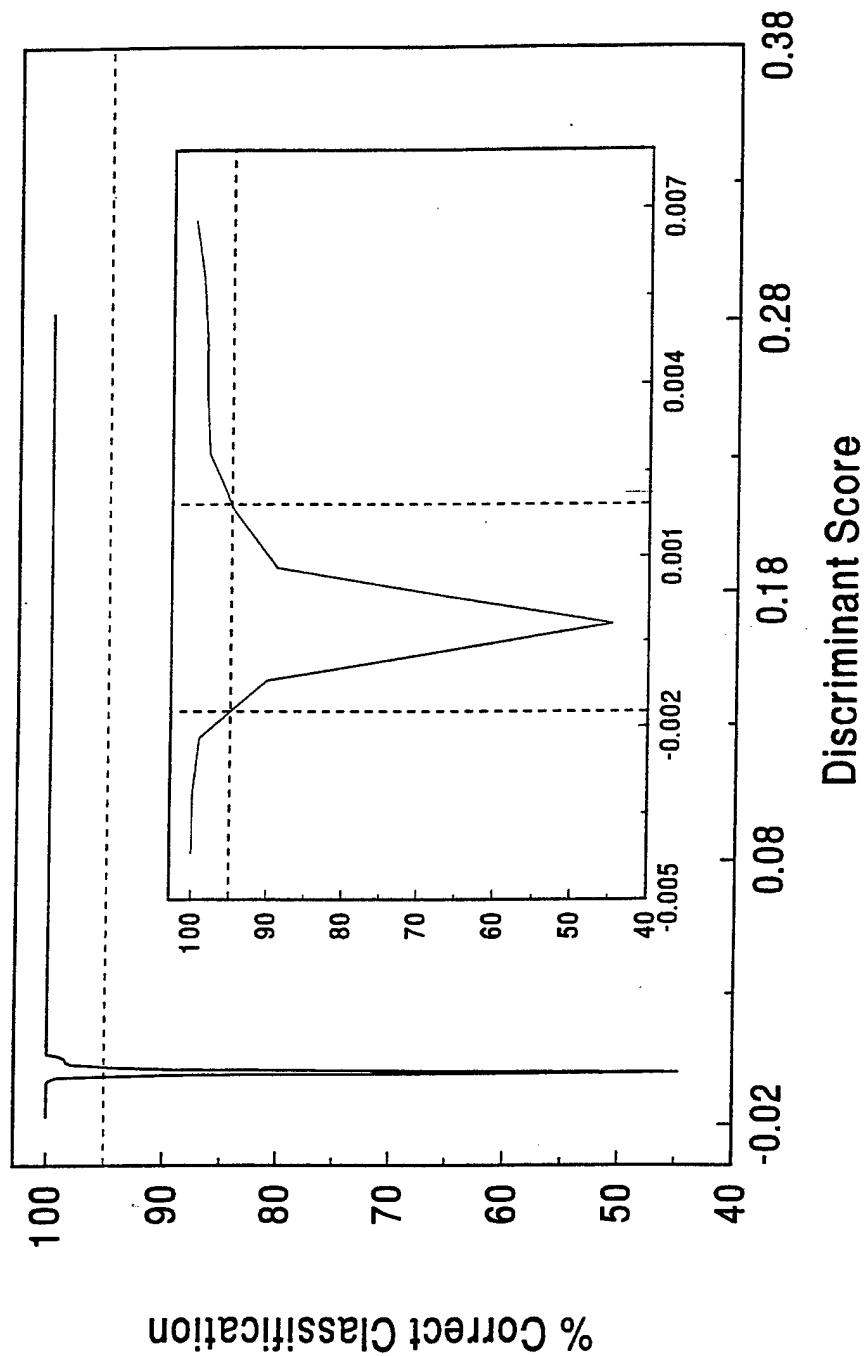
Figure 14. The percentage of correct classifications is plotted vs. discriminant score. For the 60000 interferograms in the prediction set of data set B, the corresponding discriminant scores were subdivided into bins of size 0.001. The classification percentages were computed based on the occurrence of correct and incorrect classifications among the discriminant scores falling within each bin. It is observed that classification percentages exceed 95% except when discriminant scores lie between approximately -0.0017 and 0.0019.

64

the Savitzky-Golay polynomial approximation [44,45] were applied to the single-beam spectra in an effort to enhace the TCE spectral information. Each filter was applied to each of the three spectral ranges discussed previously, producing a total of six additional data sets for use in testing the pattern recognition methodology. The same PLDA training and prediction procedures described above were used with these six data sets. Table 12 includes the resulting training and prediction results. Use of the derivative filters improves the results significantly. The best results (first derivative, 800-1350 cm^{-1}) are virtually identical to those obtained with the interferogram-based procedure. The spectral-based analysis achieves a slightly better TCE classification percentage, but is slightly more susceptible to false detections.

These results suggest that filtering strategies can be used to isolate information from the single-beam spectra directly, just as analogous procedures can be used with the interferogram. This is not surprising, given the linearity of the Fourier transform. However, as discussed previously, use of the direct interferogram analysis lowers the data collection and data processing requirements of the remote sensor and may make possible the design of rugged, low-cost spectrometers that only collect a short interferogram.

Conclusions

The results presented above demonstrate that it is possible to implement a highly accurate, selective, and automated detection of TCE by passive FTIR remote sensing measurements. The interferogram-based methodology does not require an infrared background measurement. Automated detection of TCE is achieved at the instrumental limit of detection against a wide variety of infrared backgrounds, including low-angle sky backgrounds containing a myriad of atmospheric spectral features. The ability of the interferogram-based analysis to reject these interfering spectral signatures, as well as the signatures of other chemical species present, suggests that it is viable to use bandpass filters to isolate specific spectral features in the interferogram domain. Thus, only a short interferogram segment is required to achieve significant spectral selectivity. This is confirmed by comparing the interferogram-based results to the analogous results obtained in an analysis of filtered single-beam spectra.

The analyses of data sets A and B both suggested that the optimal interferogram segment for TCE detection is located greater than 100 points from the centerburst, and that filters of FWHH $\leq$ 120 cm^{-1} are best. While the analysis of data set A was inconclusive regarding the optimal filter position, the results from data set B suggest that the use of the 938 cm^{-1} band allows the lowest limit of detection to be achieved.

Finally, a significant result derived from this work is that the piecewise linear discriminant procedure can be used to define an effective limit of detection for TCE. The discriminant scores derived from the application of PLDA to filtered interferogram data clearly encode information about the strength of the TCE spectral signal. Clear evidence also exists that the discriminant boundaries coincide with the instrumental limit of detection, and that confidence levels can be assigned to the discriminant scores. This allows the sensitivity of the automated detection algorithm to be tuned in a highly flexible manner.

Table 12

Pattern Recognition Classification Results for Spectral Data

Spectral Range (cm^{-1})	Preprocessing Method	Training (%)[a]	Prediction (%)		
		TCE-actives	TCE-actives	False detection	Total
800-1000	none	81.1	76.5	0.3	92.0
800-1200	none	95.0	92.4	0.2	97.3
800-1350	none	95.4	92.2	0.2	97.3
800-1000	1st derivative[b]	95.9	92.3	0.7	97.0
800-1200	1st derivative	98.9	97.2	0.8	98.5
800-1350	1st derivative	99.0	97.1	0.7	98.6
800-1000	2nd derivative[c]	92.6	88.4	0.2	96.0
800-1200[d]	2nd derivative	98.3	96.3	0.7	98.3
800-1350	2nd derivative	99.1	97.1	0.9	98.4

[a]Due to the single-sided requirement of piecewise linear discriminants, no false detections occur in training.

[b]Computed by a 7-point quadratic Savitzky-Golay filter.[24, 25]

[c]Computed by a 7-point quadratic-cubic Savitzky-Golay filter.[24, 25]

[d]Results computed with a four-vector piecewise linear discriminant due to the fifth vector being not single-sided.

Quantitative Analysis of Sulfur Dioxide with Passive Fourier Transform Infrared Remote Sensing Interferogram Data

One FTIR remote sensing application of significant interest is the monitoring of industrial stack emissions. In this experiment, a ground-based spectrometer is fitted with telescope optics and used to view the plume from a stack against a sky background. The analyte features in the observed spectra are typically emission bands arising from the vibrational relaxation of the hot effluents. The feasibility of performing passive remote FTIR monitoring of stack plumes has been demonstrated [46,47]. Research in a number of laboratories has focussed on either the development of the FTIR instrumentation for this application [48,49] or the development of signal processing methodologies that would be compatible with modified FTIR instruments [35.50-54].

Quantitative analysis of smokestack emissions by passive FTIR remote sensing has been hindered by several factors which affect the passive measurement of an effluent plume. These include the requirement of a significant temperature difference between the analyte and the infrared background, spectral interference caused by the presence of species such as water, ozone, and carbon dioxide, the difficulty of collecting a representative background spectrum for use in processing the spectral data of the analyte, and the effects of light scattering by airborne particulates. In addition, the use of an FTIR instrument in the outdoor environment places severe demands on the ruggedness and reliability of the spectrometer hardware.

The research described here is directed to overcoming two of the limitations listed above: eliminating the need for a spectral background measurement and fostering the development of smaller, more rugged, and more automated instrumentation. This work represents a feasibility study for testing a data analysis algorithm that has the potential for addressing both of these challenges. In this work, under experimental conditions that simulate a stack emission, quantitative analysis of sulfur dioxide is performed without the use of any background measurement by direct analysis of short segments of the collected FTIR interferograms. This is accomplished by combining a preprocessing digital filtering step and a multivariate calibration technique based on partial least-squares (PLS) regression. This approach has the additional potential advantage of decreasing both the data acquisition and data processing requirements for the measurement, as well as simplifying the instrumentation requirements. If the interferogram-based analysis can be restricted to a short segment of the interferogram, a simpler (i.e., lower resolution) interferometer design can be employed. This innovation could increase the ruggedness and reliability of a passive remote sensor, as well as reduce its manufacturing cost.

Experimentation

Instrumentation. Two Midac FTIR spectrometers (model M2400 series) were employed in this investigation (Midac Corp., Irvine, CA). The spectrometers (unit serial numbers 120 and 145) were furnished with narrow-band liquid nitrogen cooled Hg:Cd:Te detectors operating over the 1250 to 850 cm^{-1} spectral region. The field of view (FOV) for each spectrometer was limited to 3 milli-radians by a telescope having a ten-inch aperture.

A heated gas cell (Model 2408-5546, International Crystal Laboratories, Garfield, NJ) was positioned between the exit aperture of this telescope and the entrance aperture of the respective spectrometer. The gas cell characteristics were a 10 cm path length, 38 mm clear aperture, and temperature control from ambient to 250 °C with an accuracy of ± 1 °C. The gas cell valve ports were modified to allow sample flow-through operation. This cell used sodium

chloride windows to contain samples of sulfur dioxide. Since cell temperatures always exceeded ambient, fogging of the windows did not occur from accumulation of atmospheric moisture. The cell assembly was mounted directly in front of the spectrometer entrance aperture with a customized cell holder (AeroSurvey Inc., Manhattan, KS). The gas cell aperture was large enough to avoid occlusion of the spectrometer FOV through the telescope.

The interferogram data were acquired with the MIDCOL software package [55] on a Dell System 486P/50 computer operating under MS-DOS (Microsoft, Inc., Redmond, WA). The 1024-point interferograms were sampled at every eighth zero-crossing of the HeNe reference laser. The maximum digitized frequency was 1975 cm^{-1}, and the transformed spectral data had a point spacing of 3.9 cm^{-1}.

Procedures. Infrared energy collected from sky backgrounds was directed through the gas cell and into the interferometer. The cell contained either pure nitrogen or a mixture of sulfur dioxide and nitrogen. Gases were continuously flowed through the gas cell at 0.1 L/min and the output was monitored with a GASMET gas analyzer system (Temet Instruments, Oy, Finland) to ensure that a stable gas concentration was achieved. The gas flow rate was maintained through the cell during a change to a new temperature to minimize the time necessary for concentration equilibration. The cell temperatures of 50, 80, 120, and 150 °C were used in this study to simulate temperatures found in power plant stack effluent plumes. The time required to achieve a constant temperature was on the order of ten minutes. Blank measurements in which the cell contained pure nitrogen were followed by the sulfur dioxide concentrations in order of increasing concentration. Measurements for all temperatures at a given concentration were made before proceeding to the next higher concentration.

Gas concentrations were controlled by nitrogen dilution of a certified sulfur dioxide calibration gas (10,100 ppm) which was produced with ± 5% accuracy (Scott Specialty Gases, Plumsteadville, PA). The various sulfur dioxide concentrations were obtained through the use of two mass controllers that metered the appropriate relative flow rates of pure nitrogen and the sulfur dioxide calibration gas into a mixing chamber. The gas mixture was directed through the gas cell and the exit port was connected to the GASMET gas analyzer. On the basis of the determinations made with the GASMET analyzer, the path averaged concentrations introduced into the 10 cm cell varied from 238.0 to 1220.0 ppm*m. Path averaged concentrations in ppm*m units are reported for compatibility with field remote sensing measurements in which the actual optical depth of the analyte cloud is unknown.

A low-angle sky background was viewed by elevating the telescope to approximately 15° above the horizon. A set of 100 interferograms was collected for each concentration and temperature condition generated with the gas cell. The concentration was determined with the GASMET analyzer at both the beginning and end of a collection of interferograms. Seven data files were collected by use of six concentrations of sulfur dioxide and a nitrogen blank for both Midac spectrometer units. The interferograms were transferred to a Silicon Graphics 4D/460 R3000 computer (Silicon Graphics, Mountain View, CA) operating under the Irix operating system (version 5.2). All interferogram analysis was performed with this system by use of software written in FORTRAN-77. Fourier filtering and multiple linear regression analysis of interferogram data relied on the use of subroutines from the IMSL library (IMSL Inc., Houston, TX).

Results And Discussion

Overview of Emission Measurements. Infrared emission spectroscopy has been used for both qualitative [46,56] and quantitative [57-60] analysis of heated samples. The difficulties involved in quantitative analysis have been discussed in detail previously [47]. In that work, an equation for calculating the concentration of an emissive sample was derived. The derivation was based on the principal assumption that for any material,

$$T(\bar{v}) + R(\bar{v}) + \epsilon(\bar{v}) = 1 \qquad (7)$$

where $T(\bar{v})$ is the transmittance, $R(\bar{v})$ is the reflectance, and $\epsilon(\bar{v})$ is the emissivity at wavenumber, $\bar{v}$. This equation is simplified by assuming that reflectance is negligible for typical gas samples. The Beer-Lambert law can then be used to obtain the relationship between $T(\bar{v})$ and concentration, thereby producing

$$c = \frac{-\log\,[1 - \epsilon(\bar{v})]}{a(\bar{v})\,b} \qquad (8)$$

where c is the concentration of the analyte, $a(\bar{v})$ is the analyte absorptivity, and b is the optical path length of the sample material.

Emissivity, $\epsilon(\bar{v})$, is the ratio of the energy emitted by a sample at wavenumber, $\bar{v}$, to the energy emitted at that wavenumber by a blackbody radiator of the same temperature. This establishes that the measured analyte spectral response will be a function of both concentration and temperature. Thus, an evaluation of the effect of temperature on the analysis is an important part of any quantitative determination based on emission measurements.

Sulfur Dioxide Emission Bands. Sulfur dioxide is a nonlinear symmetric molecule with three characteristic fundamental vibrations at 1361 ($\bar{v}_3$), 1151 ($\bar{v}_1$), and 519 ($\bar{v}_2$) cm^{-1} in the mid-infrared spectral region. These can be attributed to the asymmetric stretch, symmetric stretch, and in-plane scissoring modes of vibration [41]. Figure 15 displays a single-beam spectrum of 916.3 ppm*m sulfur dioxide at 150 °C collected while the spectrometer (unit 120) was viewing a clear blue sky background. The emission bands of sulfur dioxide arise as positive-going peaks in the single-beam spectrum. The band at 1361 cm^{-1} is clearly visible, while the band at 1151 cm^{-1} is barely discernible above the baseline. While the band at 1361 cm^{-1} is the most visible in this spectrum, this band is often not observed in the passive FTIR spectra of an actual smokestack plume because of atmospheric attenuation due to the presence of strongly absorbing water molecules. In this work, the analysis will focus on each band individually, and the results will be compared.

Digital Filter Generation and Operation. As described previously, our laboratory has developed a series of general-purpose signal processing techniques for direct qualitative analysis of passive FTIR remote sensing interferogram data [8,9,35,54]. A combination of time-varying finite impulse response (FIR) digital filters and pattern recognition (piecewise linear discriminant analysis) methods are used to isolate the information pertaining to an analyte of interest from that corresponding to interferents or the spectral background. The digital filter

69
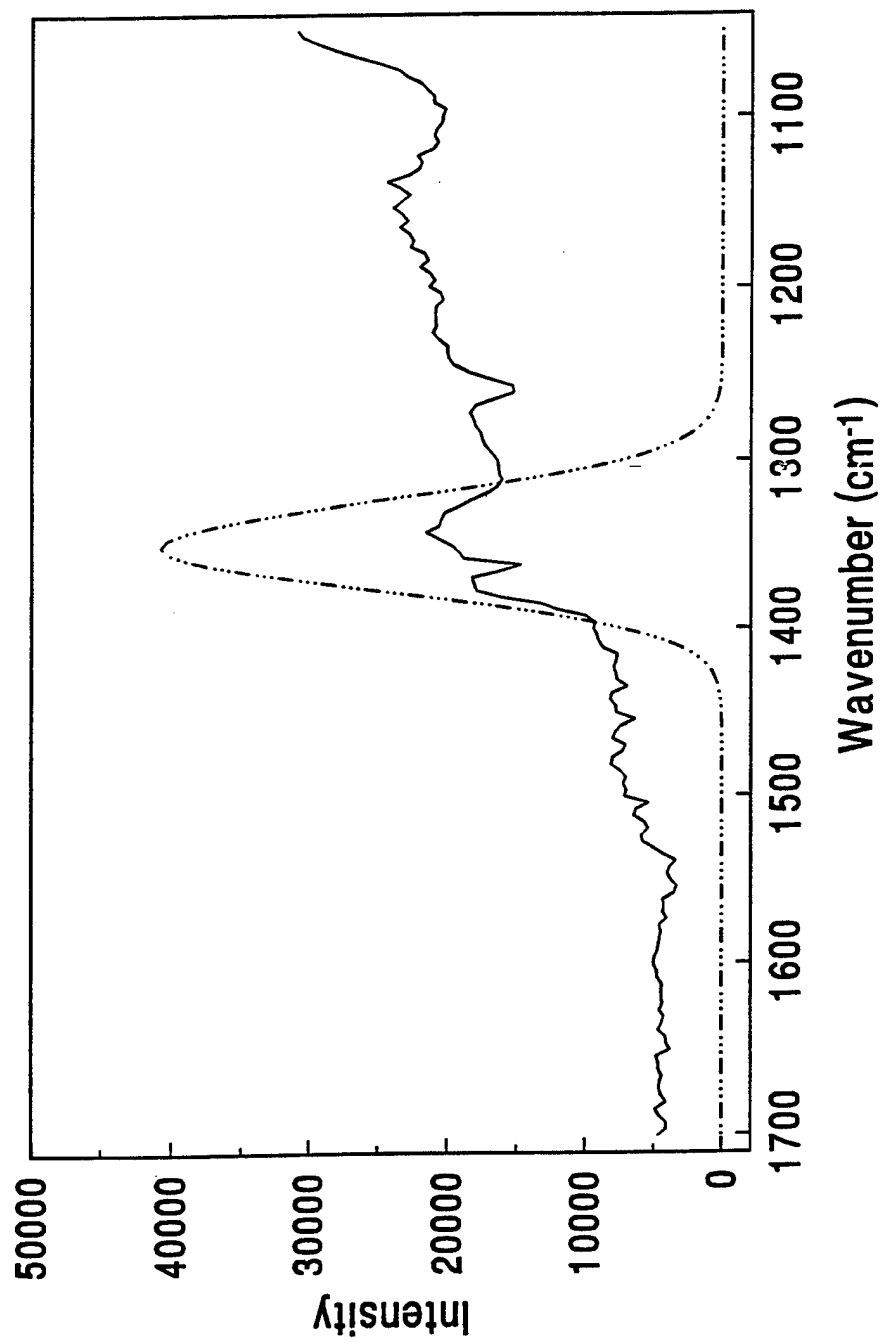
Figure 15. Single-beam spectrum of 916.3 ppm*m sulfur dioxide collected at 150 °C. The dotted profile depicts the positioning of a Gaussian-shaped bandpass digital filter designed to isolate frequencies corresponding to the sulfur dioxide emission band at 1361 cm^{-1}.

bandpass is designed to coincide with the modulated interferogram frequencies corresponding to the infrared frequencies associated with a spectral band of the analyte. The action of the digital filter is to suppress any frequency information lying outside the filter bandpass, and is therefore analogous to a background subtraction in that it removes the unwanted spectral background features before the application of the pattern recognition procedure.

The work presented here combines digital filtering and PLS regression methods to demonstrate the feasibility of performing interferogram-based quantitative analysis of sulfur dioxide using controlled field FTIR data. Successful use of digital filtering and a univariate calibration procedure with filtered interferogram data of benzene and nitrobenzene of varying concentrations has been demonstrated [29]. In that work, an approximte linear relationship between the concentration of the anayte and the intensity of the filtered interferogram was established. One of the key assumptions in this derivation was that variations in the intensity of the filtered interferogram were caused by changes in analyte concentration only. The use of the sulfur dioxide data set violates this assumption because the different temperatures contribute additional variation in the interferogram intensity. It was anticipated that the use of PLS regression would help overcome this problem.

A scheme similar to the one reported here has been successfully tested with aqueous glucose solutions which featured significant overlap between the absorption bands of glucose and water [61]. Success has also been reported with the use of a more complicated data matrix of glucose, triacetin, and bovine serum albumin mixtures of various concentrations spanning clinically relevant ranges [62]. In that study, glucose-dependent information was extracted from interferogram data by use of a combination of multiple bandpass digital filters and PLS regression.

Our data analysis procedure begins with the application of bandpass digital filters directly to short segments of the collected FTIR interferograms suspected to contain the signature of the target analyte. The design of the time-domain filter is based on the knowledge of the spectral characteristics of the target compound. For example, in this work, the filters were designed to pass the interferogram frequencies corresponding to the sulfur dioxide emission band at either 1151 or 1361 cm^{-1}. The purpose of the filtering operation on the interferogram is to provide frequency selectivity for a particular spectral band that is characteristic of the target analyte. The dotted trace in Figure 15 depicts the positioning of a Gaussian-shaped digital filter bandpass designed to isolate interferogram information pertaining to the frequencies of the sulfur dioxide band at 1361 cm^{-1}. In this work, the time-varying FIR digital filtering technique developed in our laboratory was used [8].

Figure 16 is a frequency response function of a typical filter centered at 1360 cm^{-1} and with a full width at half-height (FWHH) value of approximately 122 cm^{-1}. The units on the y-axis are attenuation in decibels (dB). Attenuation in this context means suppression of frequencies that lie outside the region specified by the filter passband.

Temperature Effects. The infrared emission radiation observed during a passive FTIR remote sensing experiment is a composite signal of emissions contributed from the target analyte, the spectrometer (i.e., cell, interferometer optics, and detector), other species in the analyte cloud and along the optical path, and the background atmospheric emissions.

For the data employed here, the variation in the IR emission is a function of the concentration and temperature of sulfur dioxide in the cell and the changes in the sky background and intervening atmosphere observed through the cell. As described by the

71
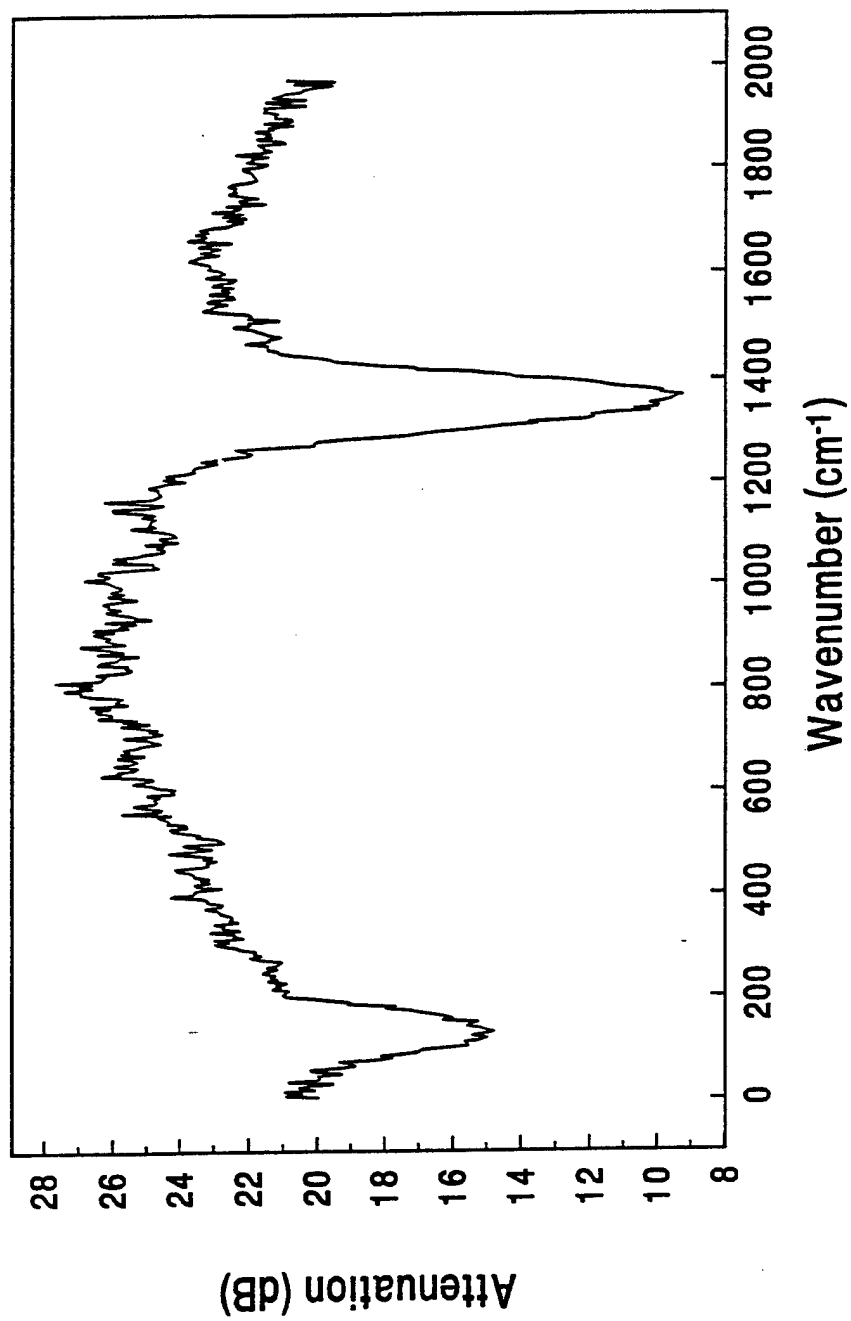
Figure 16. Frequency response function of a time-varying FIR filter consisting of an average of 14 filter coefficients. The response of the filter is plotted as attenuation in dB.

Boltzmann distribution function, temperature exerts a significant effect upon the ratio between the number of molecules in the excited and ground vibrational energy levels of a heated sample. In typical remote sensing applications, the exit temperature of stack plumes is approximately 400 K [63]. However, since the area of the analyte cloud being monitored is displaced from the top exit of the smokestack, temperature and concentration gradients of the exit gases are anticipated.

The effects of varying sulfur dioxide concentration and temperature are pictorially depicted in Figures 17A and B. Figure 17A displays passive difference spectra of three different concentrations of sulfur dioxide collected at 150 °C. A background spectrum of nitrogen collected at 150 °C was used to subtract the influence of the background from each of the spectra. As expected, an increase in the concentration of the sample is accompanied by an increase in band heights of the 1361 and 1151 cm^{-1} bands of sulfur dioxide. Figure 17B is a plot of four passive difference spectra of approximately constant sulfur dioxide concentrations, but collected at 50, 80, 120, and 150 °C. Background spectra of nitrogen collected at these four temperatures were used to subtract the effects of the background emissions from the spectra collected at the respective temperatures. It is evident that an increase in cell temperature leads to an increased intensity of emission. A nonlinear relationship between band intensity and temperature is hypothesized, as would be expected if the sample molecules obey the Boltzmann distribution. The negative intensity values in some regions of the spectra are a result of a mismatch between the analyte and background spectra used.

The digital filtering technique discussed above is designed to eliminate a majority of the background emission, but is not sufficiently selective for non-analyte signals that are heavily overlapped with the analyte signal. For this reason, PLS regression was used to help account for the various analyte and non-analyte contributions.

Assembly of Data Sets. As noted previously, during the collection of the data, two concentrations of sulfur dioxide were recorded per sub-file. Each data file contained four sub-files of 100 interferograms each, corresponding to data collected at each of the four temperature settings. This produced a total of eight sulfur dioxide concentration readings per data file. It was assumed that the two sulfur dioxide concentration readings per sub-file corresponded to the first nine and the last nine interferograms in the sub-file. This produced an average of eight samples with nine replicate interferograms each per data file.

Three data sets were assembled with interferograms collected with different combinations of cell temperatures. Each data set was constructed by randomly assigning the samples to calibration, monitoring, and prediction sets. For this assignment, a sample was defined as the nine replicate interferograms corresponding to a GASMET concentration reading.

In all studies involving the three data sets, the calibration models constructed with the use of the calibration set were evaluated with the use of the monitoring set during the optimization of the experimental parameters. The best model was the one with the lowest standard error of monitoring (SEM). SEM is defined as

$$SEM = \sqrt{\frac{\sum_{i=1}^{n_m} [c_{m,i} - \hat{c}_{m,i}]^2}{n_m}} \tag{9}$$
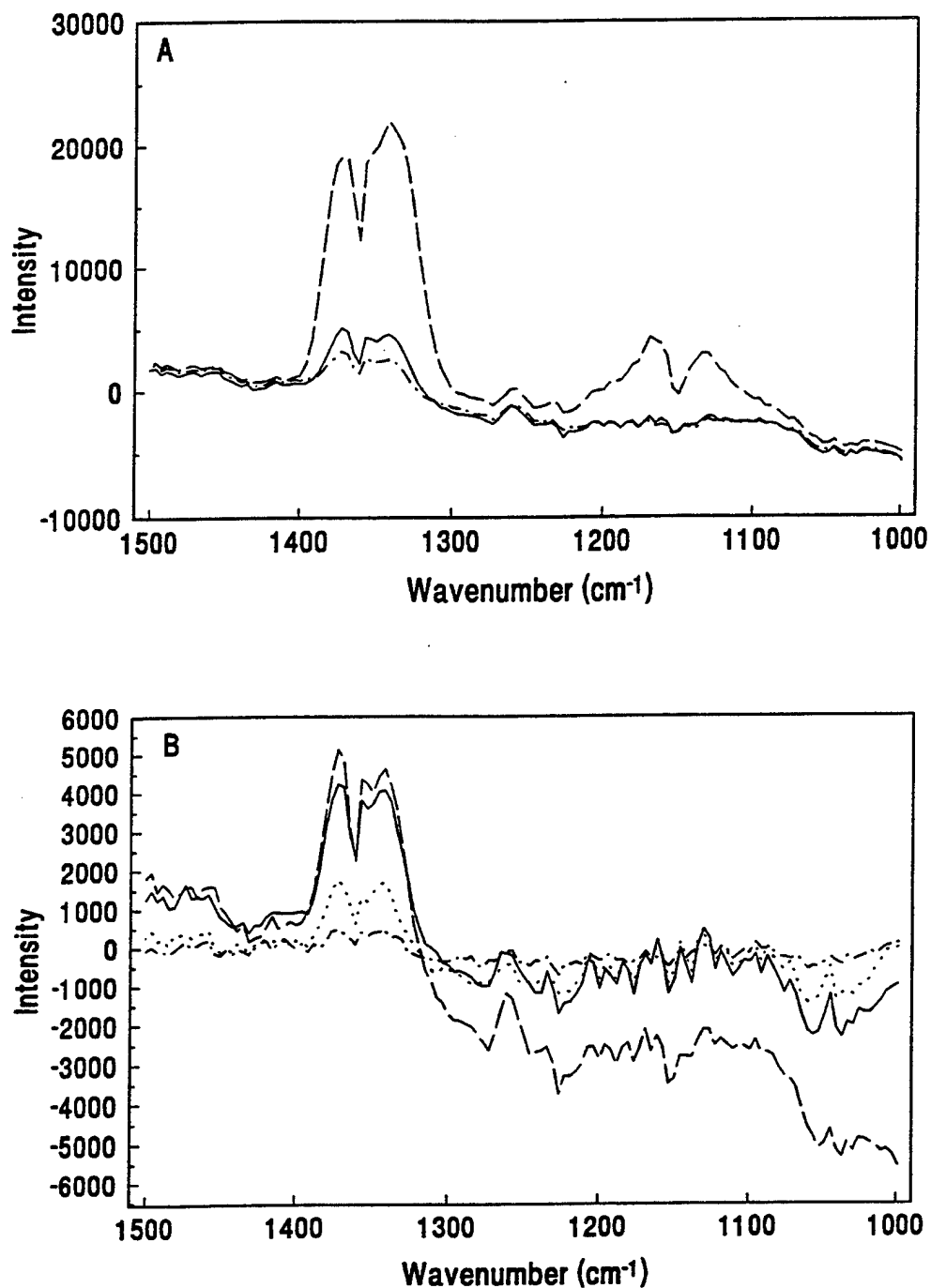
Figure 17. Passive difference spectra. (A) 1163.5 ppm*m (dashed line), 635.5 ppm*m (solid line), and 373.9 ppm*m (chain-dashed line) sulfur dioxide collected at 150 °C. (B) 635.5 ppm*m sulfur dioxide at 150 °C (dashed line), 642.9 ppm*m sulfur dioxide at 120 °C (solid line), 646.8 ppm*m sulfur dioxide at 80 °C (dotted line), and 644.9 ppm*m sulfur dioxide at 50 °C (chain-dashed line).

where n_m is the number of interferograms in the monitoring set, $c_{m,i}$ is the actual sulfur dioxide concentration associated with the ith interferogram in the monitoring set, and $\hat{c}_{m,i}$ is the corresponding sulfur dioxide concentration predicted by the model.

Once the optimal parameters were realized, the calibration and monitoring sets were combined and used to build the final calibration model. This model was tested with the use of the prediction set that had been set aside and not used during the optimization step. This way, an independent validation set was ensured within the limits of the experiment. The standard error of calibration (SEC) and standard error of prediction (SEP) were computed for the final optimal model. SEC and SEP are defined similarly to SEM, with the exception that SEC is adjusted for the loss of degrees of freedom corresponding to the number of estimated regression coefficients in the calibration model.

150 °C Data. A total of 24 samples (216 interferograms) were present in this data set. The data were randomly apportioned into 16 calibration samples (144 interferograms), 4 monitoring samples (36 interferograms), and 4 prediction samples (36 interferograms). The range of path averaged concentrations in this data set was 251.1 to 1163.5 ppm*m.

120 and 150 °C Data. There were 48 samples (432 interferograms) in this data set. The data were partitioned into 30 calibration samples (270 interferograms), 8 monitoring samples (72 interferograms), and 10 prediction samples (90 interferograms). The range of path averaged concentrations in this data set was 242.7 to 1220.0 ppm*m.

Full Data. The full data set consisted of 93 samples (837 interferograms) collected at temperatures 50, 80, 120, and 150 °C. Of these, 54 samples (486 interferograms) were from unit 120 and 39 samples (351 interferograms) were from unit 145. This data set was randomly assigned into 52 calibration samples (468 interferograms), 18 monitoring samples (162 interferograms), and 23 prediction samples (207 interferograms). The range of path averaged concentrations in the full data set was 238.0 to 1220.0 ppm*m. A complete summary of the three data sets is shown in Table 13.

Analysis of Sulfur Dioxide Band at 1361 cm⁻¹. A total of 26 digital bandpass filters centered between 1357 and 1363 cm⁻¹, each with FWHH values between 82 and 180 cm⁻¹, were designed. The filter design requires experimental data [8], and was implemented with 819 nitrogen background interferograms. Calibration models were constructed with the use of interferogram segments filtered with each of these filters. The interferogram segments studied resided between points 50 and 300, relative to the centerburst, and were 100, 150, 200, and 250 points in length. For a given number of interferogram points, the starting and stopping points were incremented by 50 points until the entire range of points 50-300 was studied. The model sizes investigated were 1 to 12 PLS factors. The optimization experiments detailed above were performed for (1) the 150 °C data, (2) the combined 120 and 150 °C data, and (3) the full data set collected at the four temperature settings. The results are tabulated in Table 14.

Analysis of Sulfur Dioxide Band at 1151 cm⁻¹. With the use of 819 nitrogen background interferograms, a total of 24 bandpass digital filters were generated. The filters were centered between 1149 and 1155 cm⁻¹, and had FWHH values that fell between 120 and 230 cm⁻¹. The rest of the optimization experiments were identical to those performed above with the 1361 cm⁻¹ band. The results are summarized in Table 15. With the exception of model size, these results are similar to those performed with the use of the more intense 1361 cm⁻¹ band. The PLS procedure requires more factors to extract information from the less intense band.

Table 13

Partitioning of Data Sets

Data Type	Calibration subset	Monitoring set	Prediction set	Total
150 °C	16[a] (144 intfgs)	4[a] (36 intfgs)	4[a] (36 intfgs)	24[a] (216 intfgs)
120 & 150 °C	30 (270 intfgs)	8 (72 intfgs)	10 (90 intfgs)	48 (432 intfgs)
All temps	52 (468 intfgs)	18 (162 intfgs)	23 (207 intfgs)	93 (837 intfgs)

[a]Number of samples defined by individual concentration measurements. The number of replicate interferograms corresponding to these samples is indicated in parentheses.

Table 14

Analysis of Sulfur Dioxide Band at 1361 cm^{-1}

Data Type	Intfg. Segment Pts[a]	PLS Factors	SEC (ppm*m)	SEP (ppm*m)	R^2 (%)
(a) Partitioned Data Sets					
150 °C	50 - 300	5	21.59	51.51 (4.4%)[b]	99.35
120 & 150 °C	50 - 150	10	24.88	52.73 (4.3%)	99.10
All temps	50 - 150	10	80.85	88.23 (7.2%)	90.08
(b) Cross-Validation Prediction					
150 °C	50 - 300	5	25.37[c]	38.67 (3.3%)	99.09[d]
120 & 150 °C	50 - 150	10	28.01	39.84 (3.3%)	98.93
All temps	50 - 150	10	81.57	95.01 (7.8%)	90.01

[a]Relative to interferogram centerburst.

[b]SEP expressed as a percentage of the maximum path averaged concentration in the data set.

[c]Pooled value computed across the set of calibration models used for the cross-validation predictions.

[d]Average value computed across the set of calibration models used for the cross-validation predictions.

Table 15

Analysis of Sulfur Dioxide Band at 1151 cm^{-1}

Data Type	Intfg. Segment Pts[a]	PLS Factors	SEC (ppm*m)	SEP (ppm*m)	R^2 (%)
(a) Partitioned Data Sets					
150 °C	100 - 300	7	19.32	49.70 (4.3%)[b]	99.48
120 & 150 °C	50 - 150	12	27.91	48.46 (4.0%)	98.87
All temps	50 - 250	12	75.00	98.93 (8.1%)	91.49
(b) Cross-Validation Prediction					
150 °C	100 - 300	7	23.01[c]	40.04 (3.4%)	99.26[d]
120 & 150 °C	50 - 150	12	29.90	44.58 (3.7%)	98.79
All temps	50 - 250	12	77.46	103.03 (8.4%)	91.01

[a]Relative to interferogram centerburst.

[b]SEP expressed as a percentage of the maximum path averaged concentration in the data set.

[c]Pooled value computed across the set of calibration models used for the cross-validation predictions.

[d]Average value computed across the set of calibration models used for the cross-validation predictions.

Cross-Validation Prediction. In order to validate the results obtained from the studies above, the alternative calibration method of cross-validation was used. This method is most commonly used in applications where the number of available samples is limited. In such situations dividing the data into calibration and prediction sets would lead to data subsets that do not encode the total span of variation present in the full data set, resulting in poorly predicting calibration models. These models predict poorly because they underestimate the errors to be expected in a true unknown sample. Therefore, the error in this cross-validated prediction should be more representative of what one would obtain with an independent set of unknown samples that have been represented adequately by the calibration samples.

The cross-validation method used was the leave-one-out type [64]. Given a set of n calibration samples, a calibration model was built with $n - 1$ samples, and using this calibration model, the concentration of the sample left out was predicted. This process was repeated n times until each sample has been left out and predicted once. In this procedure, the nine replicate interferograms corresponding to each sample were all left out and predicted as a group. Ideally, this procedure should have been used when optimizing the experimental variables but is prohibitively time consuming, especially with a relatively large data set such as the full data set above. Therefore, in this work, cross-validation was performed with the optimal parameters already realized above. The results from this study are tabulated in Tables 14 and 15.

Evaluation of Results. For the 150 °C data, using the sulfur dioxide band at 1361 cm^{-1}, the best calibration model had R^2, SEC, and SEP values of 99.09%, 25.37, and 38.67 ppm*m, respectively. This was a five-factor model realized using a 250-point interferogram segment, located between points 50 and 300, relative to the centerburst. The filter was centered at 1359.6 cm^{-1} and the FWHH value was 122 cm^{-1}. The combined 120 and 150 °C data required a 10-factor model along with a 100-point interferogram segment located between points 50 and 150 to produce equivalent results. The filter used was centered at 1363 cm^{-1} and had a FWHH value of 142 cm^{-1}. This model had an R^2 of 98.93%, a SEC of 28.01 ppm*m, and an SEP of 39.84 ppm*m. As the number of samples from the separate instruments increased, additional sample and instrumental variations are introduced into the calibration model, requiring additional PLS factors to account for them. Figure 18 displays two correlation plots corresponding to the best results from the (A) 150 °C and (B) combined 120 and 150 °C data. Both plots show an excellent correlation between the estimated and actual sulfur dioxide concentrations, confirming the suitable choice of a linear model. The prediction samples (closed triangles) fall within the spread of the calibration samples (open circles).

Figure 19 shows a correlation plot generated from the best calibration results of the full data set. This data set includes all samples from the two spectrometers collected at 50, 80, 120, and 150 °C. The 10-factor model realized with the use of a 100-point interferogram segment located between points 50 and 150 had an R^2 of 90.01%, a SEC of 81.57 ppm*m, and a SEP of 95.01 ppm*m. From the correlation plot, it is evident that the data analysis procedure used here cannot account for variations in infrared intensities caused by temperature differences of up to 100 °C. However, the prediction samples (solid triangles) still fall within the spread of the calibration samples (open circles).

The PLS calibration model is based on the assumption of a linear relationship between the intensities of filtered interferogram points and the concentration of sulfur dioxide. This assumption is not necessarily valid because unlike in absorption spectroscopy, there is no linear equation that relates emissivity to analyte concentration in emission spectroscopy. Also in order for eqn. 8 to be applicable in calculating concentrations, eqn. 7 must be true to a very good

78
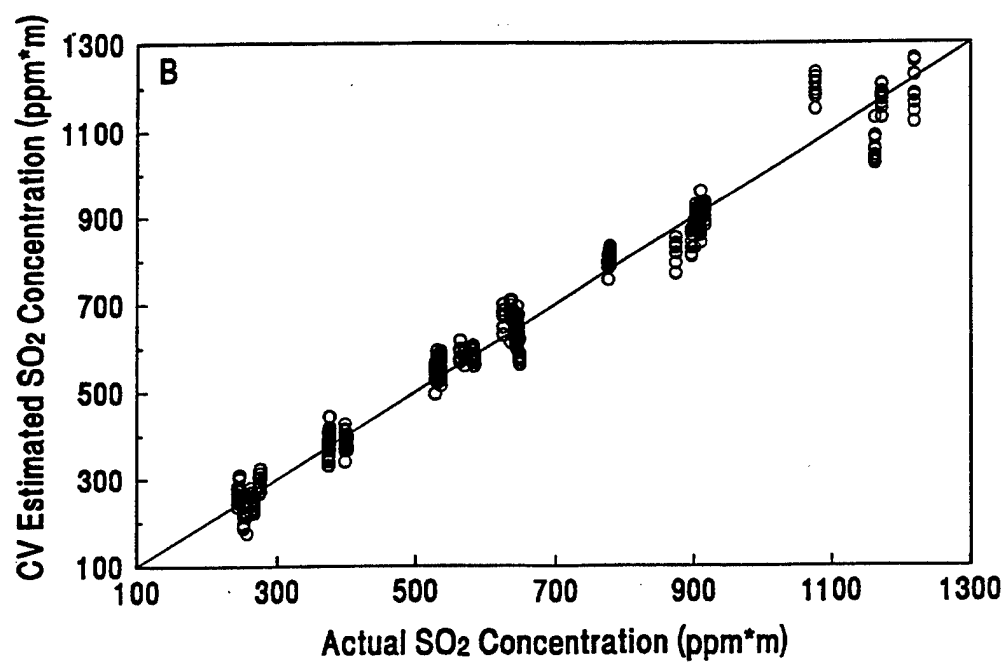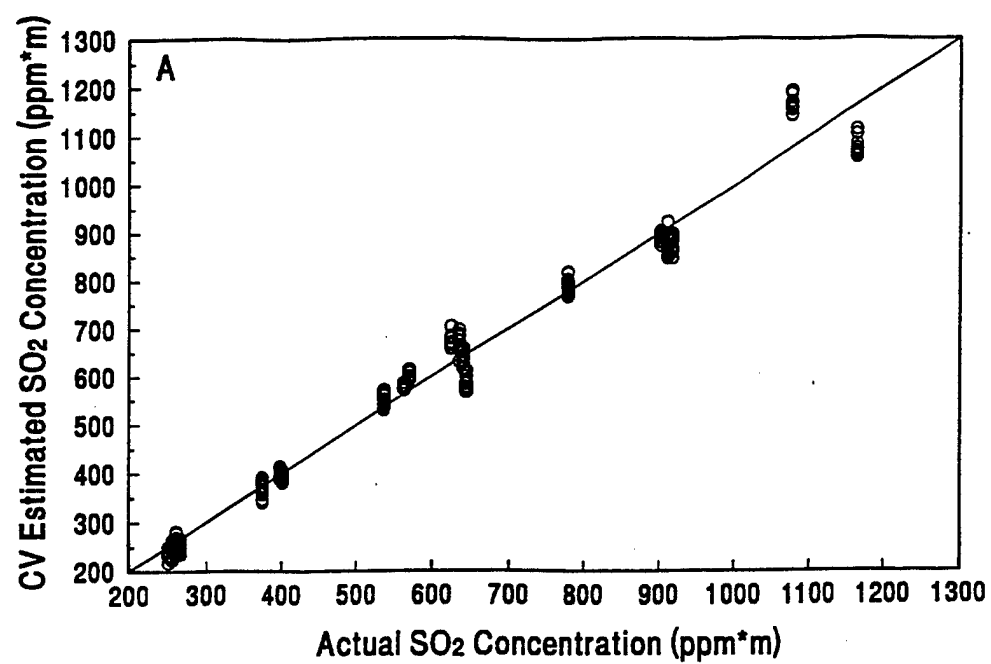
Figure 18. Correlation plots of cross-validated estimated vs. actual sulfur dioxide concentrations. (A) 150 °C data and (B) combined 120 and 150 °C data.
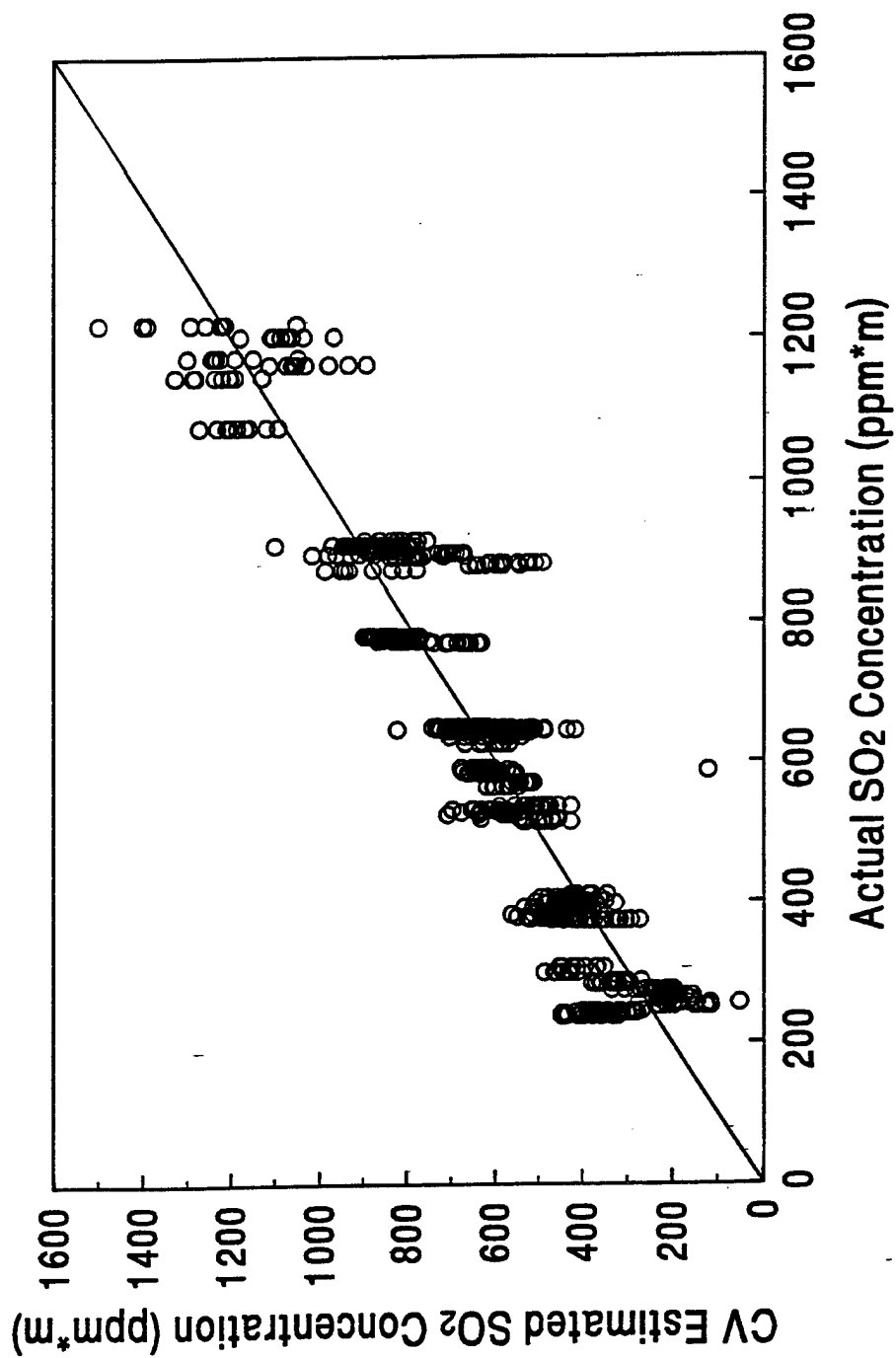
Figure 19. Correlation plots of cross-validated estimated vs. actual sulfur dioxide concentrations corresponding to data collected at 50, 80, 120, and 150 °C.

degree of approximation. Residual plots derived from the calibration models can be used to detect a lack of fit to the linear model. Figure 20 shows residual plots from the cross-validation results corresponding to (A) 150 °C data and (B) combined 120 and 150 °C. From the two plots, it is evident that the calibration and prediction of high concentrations of sulfur dioxide poses a significant challenge to the calibration model. This could be attributed to the use of a linear algorithm for modeling a phenomenon that is nonlinear at high concentrations. Another compelling evidence of nonlinearity is the number of PLS factors required to model the variation present in these data sets. A similar pattern was observed in the residual plot corresponding to the best calibration results of the full data set. The use of a formal nonlinear modeling technique such as artificial neural networks or nonlinear PLS regression is currently under investigation to address this problem.

Conclusions

This work has demonstrated that quantitative analysis of passive remote sensing FTIR data can be implemented by use of short segments of bandpass filtered interferograms. The results from this work confirm that a linear model can be used to approximate the relationship between filtered interferogram intensities and the concentrations of sulfur dioxide emissions for data collected with temperature differences $\leq$ 30 °C. This is a significant result because the technique can to some significant degree correct for the temperature differences between the molecules of the target analyte in the area of the cloud being monitored. Although, for the full data set the proposed method could not correct for signal variation introduced by temperature variation of 100 °C, it has been reported that within a stack diameter of the top of the stack, temperature and concentration gradients are minimal [65]. Therefore, judicious selection of the area to focus the telescope in the plume could decrease the effects of temperature variation.

The analysis was performed without the use of a background reference interferogram, thereby circumventing the virtually impossible task of collecting a "clean" non-varying background interferogram in FTIR remote sensing applications. That the analysis could be performed with a short interferogram segment in the time-domain is a very significant result because the next generation of FTIR remote sensors could be engineered around this result. The optical retardation of the moving mirror in the interferometer compartment would be minimized, thereby making the sensor potentially more rugged, reliable, and more suited for mounting on a moving vehicle or on an airborne platform. The methodology also has the potential for automation and real-time quantitative analysis of a large number of compounds in the workplace.

Calibration Transfer Results for Automated Detection of Acetone and Sulfur Hexafluoride by FTIR Remote Sensing Measurements

Two critical problems have been shown to hinder the widespread application of FTIR remote sensing methods to the monitoring of airborne pollutants. Traditionally in FTIR remote sensing, a reference background spectrum is collected and used to remove the background emission profile present in analyte spectra [1]. Simple changes in the environment such as wind or temperature often prohibit stable, reproducible reference spectra from being measured. Analyte spectra obtained in this fashion contain widely varying baselines and can be difficult to

81
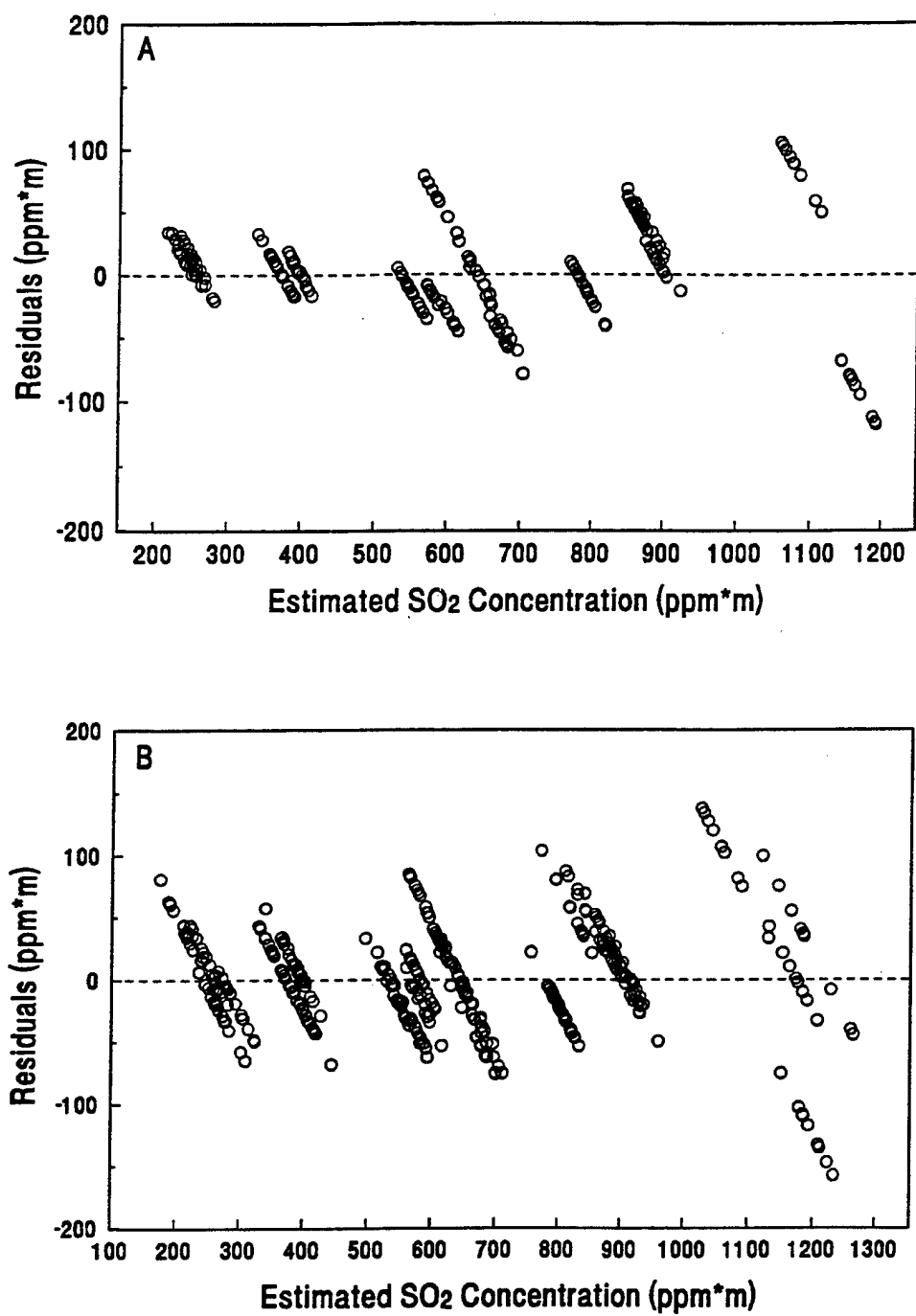
Figure 20. Residuals vs. cross-validated estimated sulfur dioxide concentration. (A) 150 °C data and (B) combined 120 and 150 °C data.

analyze. In addition, a second important background problem is the large instrument-specific signatures which make the automated analysis of data from different instruments difficult.

The interferogram-based data analysis methodology described above seeks to overcome these challenges through signal processing and pattern recognition techniques applied directly to the raw interferogram data obtained from the passive remote sensing spectrometer, avoiding the need altogether of a separate background measurement. As described previously, digital filtering steps isolate the analyte signal from the background, and pattern recognition techniques are utilized to discriminate and characterize signals which contain analyte from those which do not in an automated fashion. In the research described here, this methodology is extended with additional signal processing steps and with more strongly attenuating digital filtering techniques. It will be shown that instrument-specific background problems can be eliminated as well, allowing a successful transfer of qualitative calibration information between spectrometers.

Experimentation

Calibration transfer issues in passive remote sensing were explored by collecting laboratory acetone and sulfur hexafluoride (SF_6) interferograms on a pair of similarly configured Midac Outfielder FTIR emission spectrometers, labeled units 120 and 145 (Midac Corp., Irvine, CA). These spectrometers employed liquid nitrogen-cooled Hg:Cd:Te detectors for use in the 800-1400 cm^{-1} spectral range.

These spectrometers were interfaced to a Dell system 486P/50 IBM PC compatible computer (Dell Computer, Austin, TX) operating under MSDOS (Microsoft, Redmond, WA). Data acquisition was performed with the MIDAS software package [38]. A maximum spectral frequency of 1974.75 cm^{-1} was obtained with interferogram points being collected at every eighth zero crossing of the reference laser. Four cm^{-1} point spacing was obtained through the collection of 1024 interferogram points per scan.

A 4x4 inch extended blackbody (Model SR-80, CI Systems, Agoura, CA) provided a NIST traceable infrared source whose temperature was varied over 5 to 50 °C. The source temperature was accurate to 0.03 °C and precise to ± 0.01 °C. A sample gas cell with windows composed of low density polyethlyene (0.0005 in. thickness) was used. A thermocouple was utilized to monitor gas cell temperature. Reagent grade acetone and sulfur hexafluoride were used as analytes.

For both acetone and SF_6 experiments, data collection for units 120 and 145 was performed alternately by moving the cell and blackbody in front of each instrument in turn. For the acetone data set, interferograms were collected with blackbody temperatures from 5 to 50 °C with steps at approximately 5 °C intervals for dilution factors with water of 1 (pure acetone), 1/2, 1/4, 1/8, 1/16, 1/32, and 1/64. Between 20 and 200 interferograms were collected at each level. For the SF_6 data set, interferograms were collected over the same temperature range with similar 5 °C steps with injected analyte volumes of 0.05, 0.02, 0.1, 0.2. 0.3, 0.5, and 1.0 cc. Between 20 and 150 interferograms were acquired at each level.

The collected interferograms were Fourier transformed and the resulting single-beam spectra were ratioed to corresponding background spectra collected when no analyte was present. After converting to absorbance units, the spectra were visually inspected to ensure that the analyte signal was clearly visible above the noise. Those which did not meet this criterion were removed from the data set. This led to the training and prediction sets for each analyte as listed in Table 16.

For data analysis, the collected data sets were transferred to a dual 180 MHZ Pentium Pro (Intel Corp., Santa Clara, CA) personal computer operating under the Linux operating system, version 2.0.14. The digital filtering and pattern recognition were performed on this system with original software written in FORTRAN 77 and C. Additional processing was performed with the aid of Matlab version 4.2c (The MathWorks, Natick, MA).

Results and Discussion

Infrared signals measured through passive FTIR remote sensing experiments consist of analyte, background, and instrument-specific features superimposed [35]. The lack of a stable background prevents the use of conventional data analysis methods such as the calculation of absorbance or difference spectra for an automated determination since they are unable to remove background and instrument features reliably through ratioing or subtraction. The purpose of the signal processing and pattern recognition steps outlined here is the extraction of analyte information and the suppression of interfering signals, thereby allowing an automated determination to be performed without the use of background measurements for ratioing or subtraction.

For analytes in this study, the features of interest are the 1216 cm^{-1} C-CO-C stretching band of acetone (49 cm^{-1} full width at half maximum (fwhm)) and the 945 cm^{-1} S-F stretching band of SF_6 (10 cm^{-1} fwhm). Figure 21 demonstrates the type of signal obtained through the calculation of absorbance spectra for interferograms collected from the laboratory acetone and SF_6 data sets. For the blackbody source temperature range covered in this study, both absorption and emission peaks were present in the data sets. Fine rotation features were absent in all spectra calculated from these data due to the 4 cm^{-1} spectral point spacing.

In order for our methodology to avoid the use of inactive backgrounds for ratioing or subtraction, signal processing and pattern recognition analysis are applied directly to the interferogram data. Direct interferogram analysis provides advantages by decomposing spectral features of different widths into different regions of the interferogram. This can be attributed to the fact that the interferogram representation of a narrow spectral feature dampens more slowly than the corresponding representation of a wide background feature. By optimal choice of the interferogram segment to use for analysis, a significant amount of background interference can be removed.

Once an optimal segment is isolated from the interferogram, digital filtering is used to enhance the analyte signal further. Time domain digital filtering involves the estimation of the convolution of the interferogram with the time domain representation of the filter frequency response function [8]. Digital filtering provides a means of extracting frequency information due to the analyte from the problematic background frequencies while allowing the methodology to utilize key advantages found in signal processing data in the interferogram (time) domain.

Two types of digital filtering were used in this study, a time-varying finite impulse response matrix filter (FIRM) developed previously in our laboratory, and a standard FIR filter [8]. FIRM filters sacrifice attenuation but offer high computational efficiency by having fewer coefficients. During filter generation, coefficients deemed statistically insignificant in the estimation of the convolution sum can be discarded. Standard FIR filters were calculated

Table 16

Partition of Acetone and SF$_6$ Data Sets

	Acetone		SF$_6$	
	Unit 120	Unit 145	Unit 120	Unit 145
Training	3170[a] (8782[b])	3239 (8284)	2640 (3940)	2041 (3942)
Prediction	2190 (8202)	2292 (7753)	2320 (3773)	2134 (3776)
Collected	6810 (16984)	6810 (16037)	8273 (7713)	8275 (7718)

[a]Analyte active.

[b]Analyte inactive.

Table 17

FIRM Filter Parameters

Variable	SF$_6$	Acetone
Filter bandpass width (fwhm)	36.4[a] (81[b]), 45.4 (110), 54.5 (125), 63.6 (146), 72.7 (165) cm^{-1}	45.4 (85), 54.5 (103), 63.6 (150), 72.7 (167), 81.9 (201) cm^{-1}
Interferogram segment location[c]	75, 100, 125, 150, 175	50, 75, 100, 125, 150

[a]Specified fwhm during filter generation.

[b]Measured fwhm.

[c]Relative to interferogram centerburst.
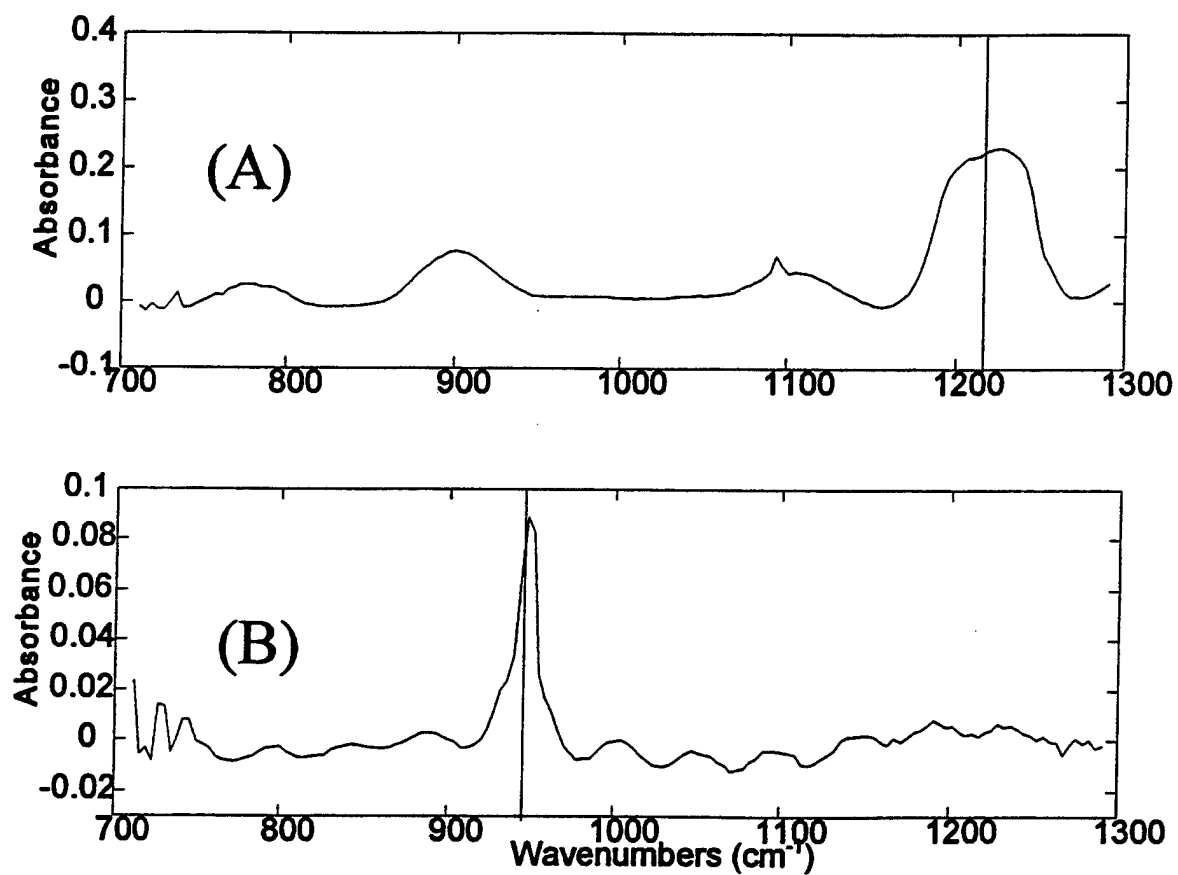
Figure 21. SF$_6$ and acetone absorbance spectra collected on Midac unit 120 under laboratory conditions. (A) Pure acetone spectrum at a blackbody temperature of 50 °C. The line at 1216 cm^{-1} highlights the acetone peak. (B) 1 cc SF$_6$ at a blackbody temperature of 50 °C. The SF$_6$ band at 945 cm^{-1} is highlighted.

through the Remez exchange algorithm and provide exceptional out-of-band attenuation; however they contain nearly an order of magnitude more coefficients. Figure 22 shows frequency response plots for a representative SF_6 FIRM filter as well as several FIR filters utilized in this study. FIR filtering allows a closer approximation of the desired passband width to be attained. However, the FIRM filter attains approximately 25 decibels (dB) of attenuation with an average of 22 filter coefficients, whereas the FIR filters all contain 200 filter coefficients.

After filtering, a reliable pattern recognition step is required in the analysis to determine the presence or absence of analyte signal in the filtered data. Due to its high performance and simplicity in configuration, the nonlinear pattern recognition technique utilized for this methodology was piecewise linear discriminant analysis (PLDA). PLDA attempts to optimize the location of linear separating surfaces, termed discriminants, which divide the data space into analyte-active and inactive categories [9,13].

As described above, previous work has demonstrated the most efficient means of optimizing the experimental parameters of FIRM filter passband center and width, interferogram segment starting position and length, as well as those of the PLDA pattern recognition algorithm. Using this protocol, and a subset of the overall experimental design used previously, FIRM filters were created with the same characteristics for SF_6. Acetone FIRM filters were also created, but with segment location and filter passband center optimized for its 1216 cm^{-1} peak. These filters were utilized to examine training and prediction as well as calibration transfer issues for acetone and SF_6. Table 17 summarizes the FIRM filter parameters used. Two values are indicated for FIRM filter width. The first is the width supplied to the filter generation algorithm, while the second width is the fwhm measured from the actual frequency response of the generated filter. Absolute values of the training and prediction interferograms were used in order to make the data space more robust for calibration transfer, and Forman phase correction was utilized. In all cases Midac unit 120 was used as a primary instrument, meaning that its interferograms were used during filter generation, as well as during pattern recognition training. Midac unit 145 was used as a secondary instrument to test calibration transfer. No unit 145 interferograms were included during training.

Results for FIRM filtering experiments from data collected on unit 120, and then utilized for both training and prediction were between 88.45 and 99.93% for both SF_6 and acetone. These results demonstrate that FIRM filtering performs well for same-instrument prediction for both analytes, as has been shown in the past. However, once these same discriminants were applied to data from a secondary instrument (unit 145), cross-prediction results decreased as seen in Figures 23 and 24, particularly for SF_6. At ~40 cm^{-1}, the acetone spectral feature at 1216 cm^{-1} is approximately four times wider than SF_6. The typical FIRM passband more closely approximates the wider acetone peak, but lets a great deal of background information through for the narrow SF_6 peaks. Cross-prediction results appear to improve as acetone FIRM filter passband widths increase, however no clear trend is evident for the optimal segment location.

Although an extensive experimental study has yet to be performed with FIR filter parameters similar to that done with the FIRM study, four FIR filters were generated with constant passband width and varying attenuations for both acetone and SF_6. These filters were applied to the same interferogram segment positions used in the FIRM study. Frequency responses for these four filters for SF_6 can be seen in Figure 22, with those of acetone being similar except for the passband center being located at 1216 cm^{-1}.
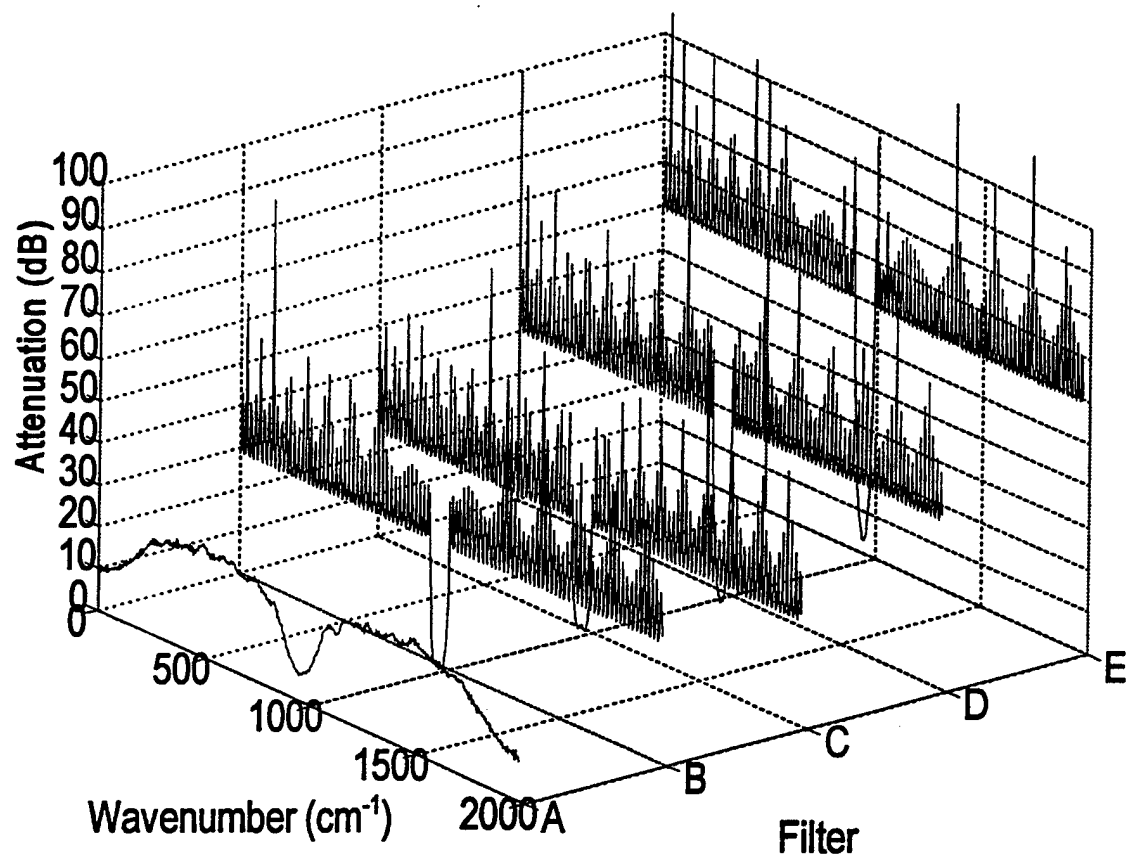
Figure 22. SF_6 FIRM and FIR filter frequency response plots demonstrating differences in atttenuation and passband width. (A) FIRM filter with fwhm ~165 cm^{-1}. (B-E) FIR filters with fixed passband width of 72 cm^{-1}.
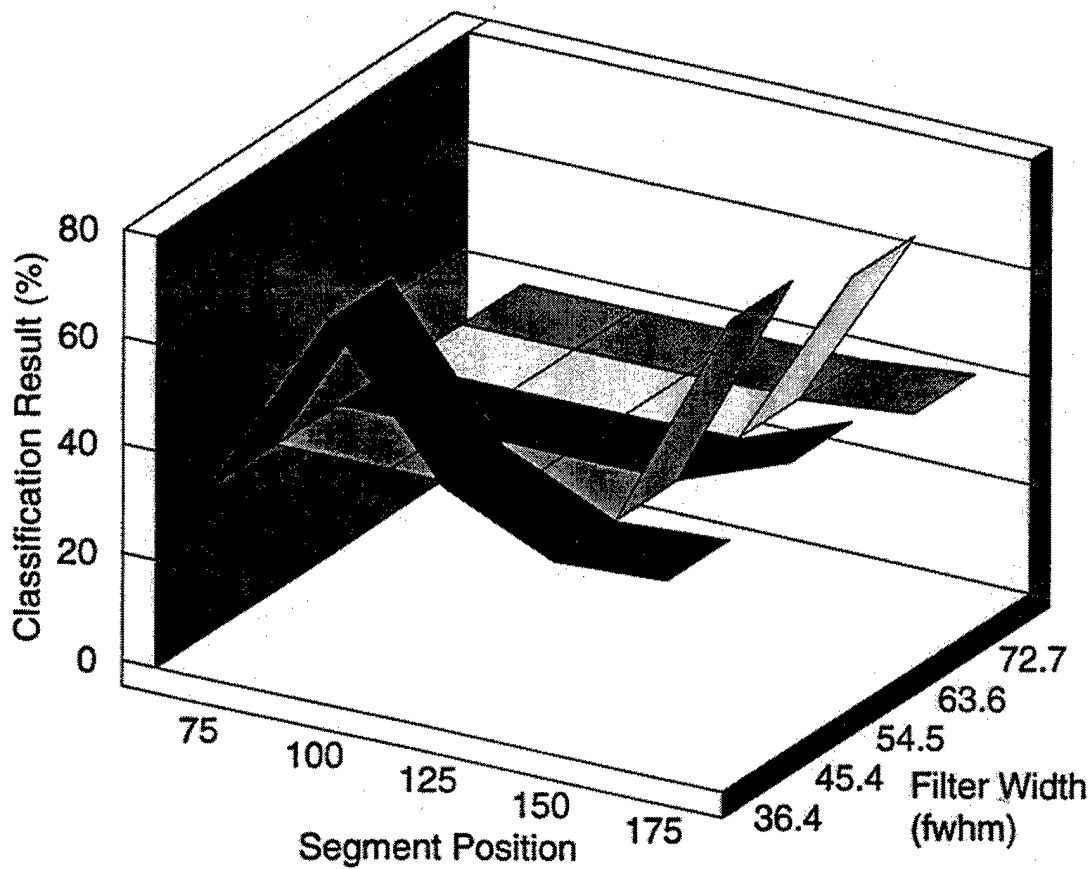
Figure 23. FIRM filtering cross-prediction results for SF_6. Midac unit 120 was used as the primary instrument in predicting the unit 145 data set.
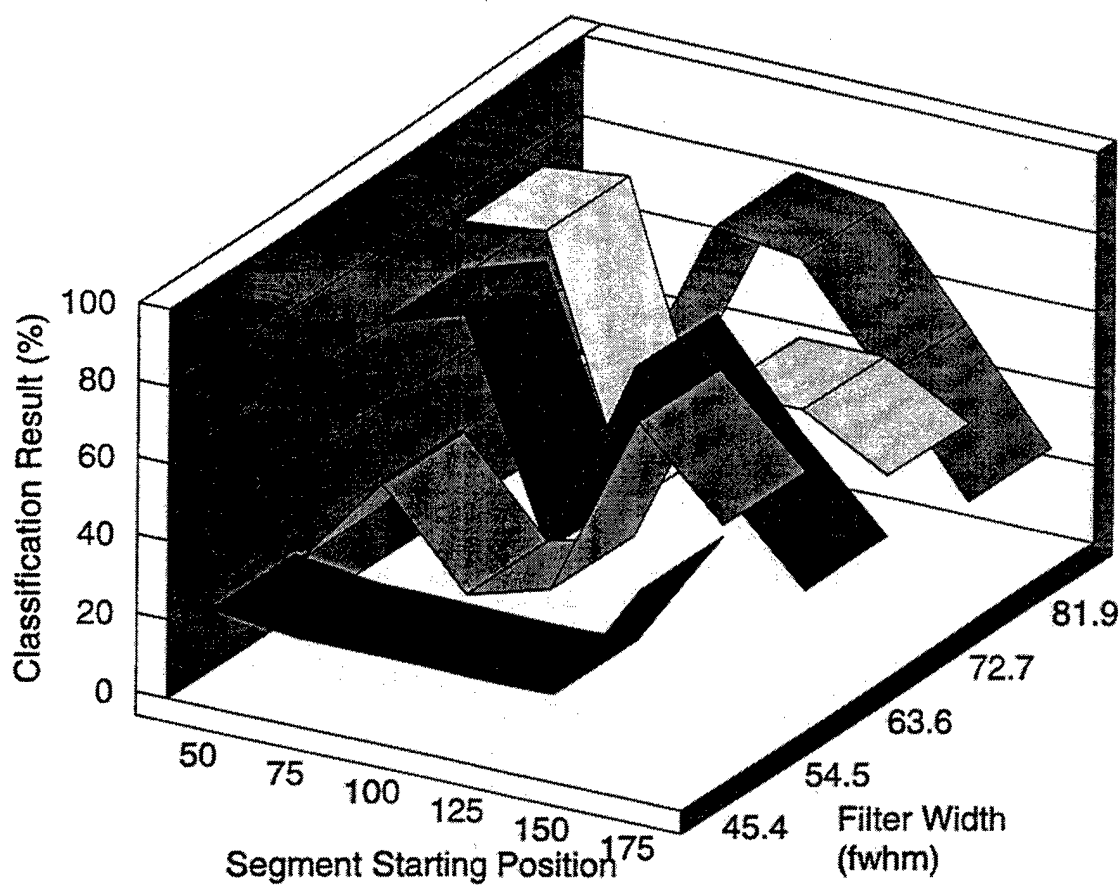
Figure 24. FIRM filtering cross-prediction results for acetone. Midac unit 120 was used as the primary instrument for predicting the unit 145 data set.

Results for same-instrument prediction for both analytes varied between 89.99 and 99.98%, and were similar to the results obtained with the FIRM filters. However, as seen in Figures 25 and 26, cross-prediction scores for both compounds were markedly improved. The acetone and SF_6 predictions are observed to improve with increasing attenuation in the stopband, with the best results being observed for attenuations above 60 dB and segments located past point 125 (relative to the centerburst).

Conclusions

While FIRM filters provide sufficient performance for training and prediction on a single instrument, FIR filters with high degrees of stopband attenuation allow a successful transfer of qualitative calibration information across data spanning two spectrometers for both acetone and SF_6 analytes.

Summary

This report described studies directed to the automated analysis of FTIR remote sensing interferogram data. The research presented here demonstrated that both qualitative and quantitative information can be extracted from short segments of digitally filtered interferograms without the need for any background or reference measurement. Through the use of experimental design techniques, an optimization protocol was devised for determining the key implementation parameters of the interferogram-based analysis. An automated compound identification algorithm was then developed for TCE and was shown to operate effectively in the presence of a wide variety of infrared backgrounds. A quantitative analysis for SO_2 was also implemented through the direct use of short interferogram segments. The TCE and SO_2 studies illustrate that it is feasible to perform both qualitative and quantitative air monitoring measurements with the interferogram-based data analysis methodology. Finally, it was shown that the analysis can be made resistant to instrument-dependent artifacts. In this way, the data analysis protocols can be developed with data from one spectrometer and then applied to data collected with a second instrument. This capability is extremely encouraging, and may make possible the large-scale implementation of automated compound detection and quantitation capabilities.
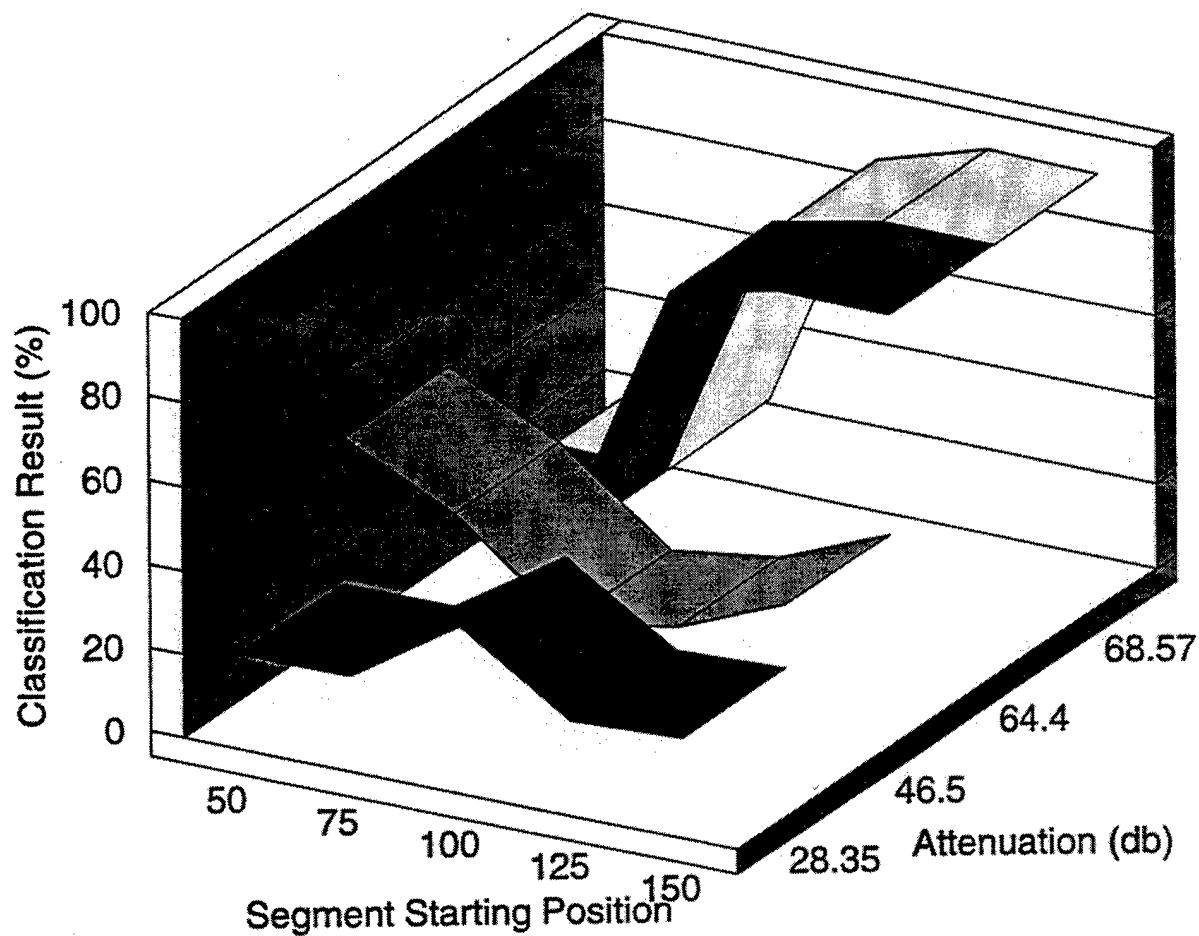
Figure 25. FIR filtering cross-prediction results for SF_6. Midac unit 120 was used as the primary instrument for predicting the unit 145 data set.
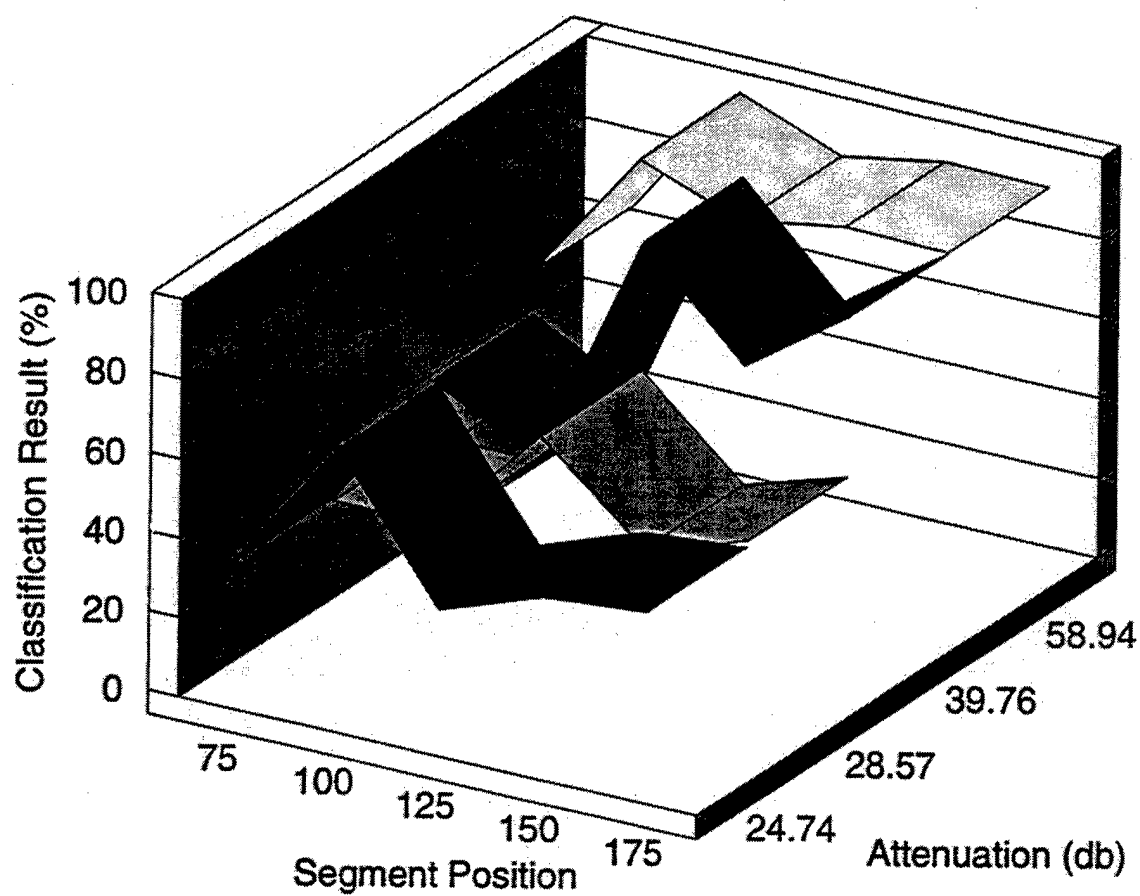
Figure 26. FIR filtering cross-prediction results for acetone. Midac unit 120 was used as the primary instrument for predicting the unit 145 data set.

Blank

References

1. Hammaker, R. M.; Fateley, W. G.; Chaffin, C. T.; Marshall, T. C.; Tucker, M. D.; Makepeace, V. D.; Poholarz, J. M. *Appl. Spectrosc.* **1993**, *47*, 1471-1475.

2. Beer, R. *Remote Sensing by Fourier Transform Spectrometry*; Wiley-Interscience: New York, 1992.

3. Small, G. W.; Carpenter, S. E. *Appl. Spectrosc.* **1992**, *46*, 49-59.

4. Small, G. W.; Barber, A. S. *Chemom. Intell. Lab. Syst.*, **1992**, *15,* 203-217.

5. Carpenter, S. E.; Small, G. W. *Appl. Spectrosc.* **1992**, *46*, 1790-1798.

6. Kaltenbach, T. F.; Small, G. W. *Anal. Chim. Acta* **1993**, *279*, 309-322.

7. Carpenter, S. E.; Small, G. W. *Anal. Chim. Acta* **1991**, *249*, 305-321.

8. Small, G. W.; Harms, A. C.; Kroutil, R. T.; Ditillo, J. T.; Loerop, W. R. *Anal. Chem.* **1990**, *62,* 1768-1777.

9. Kaltenbach, T. F.; Small, G. W. *Anal. Chem.* **1991**, *63,* 936-944.

10. Small, G. W.; Kaltenbach, T. F.; Kroutil, R. T. *Trends Anal. Chem.*, **1991**, *10,* 149-155.

11. Bangalore, A. S.; Small, G. W.; Combs, R. J.; Knapp, R. B.; Kroutil, R. T. *Anal. Chim. Acta* **1994**, *297*, 387-403.

12. Kroutil, R. T.; Combs, R. J.; Knapp, R. B.; Small, G. W. *Appl. Spectrosc.*, **1994**, *48*, 724-733.

13. Shaffer, R. E.; Small, G. W. *Chemom. Intell. Lab. Syst.* **1996**, *32*, 95-109.

14. de Haseth, J. A.; Isenhour, T. L. *Anal. Chem.* **1977**, *49*, 1977-1981.

15. Brissey, G. M.; Henry, D. E.; Giss, G. N.; Yang, P. W.; Griffiths, P. R.; Wilkins, C. L. *Anal. Chem.* **1984**, *56*, 2002-2006.

16. Bjerga, J. M.; Small, G. W. *Anal. Chem.* **1989**, *61*, 1073-1079.

17. Monfre, S. L.; Brown, S. D. *Appl. Spectrosc.* **1992**, *46*, 1699-1710.

18. Monfre, S. L.; Brown, S. D. *Appl. Spectrosc.* **1992**, *46*, 1711-1718.

19. Shikarkhane, N. S.; Nundy, U.; Chatterjee, U. K. *Appl. Opt.* **1988**, *27*, 1636-1638.

20. Combs, R. J.; Knapp, R. B. Unclassified Technical Report CRDEC-TR-387; U.S. Army, Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, July 1992.

21. Tou, J. T.; Gonzalez, R. C. *Pattern Recognition Principles*; Addison-Wesley: Reading, MA, 1974; pp. 119-123.

22. Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*; John Wiley: New York, 1978; Chapter 7.

23. Daniel, C. *Applications of Statistics to Industrial Experimentation*; John Wiley: New York, 1976; Chapter 8.

24. Neter, J.; Wasserman, W.; Kutner, M. H. *Applied Linear Statistical Models*; Irwin: Boston, MA, 1990; Chapters 16, 22.

25. Mason, R. L.; Gunst, R. F.; Hess, J. L. *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*; John Wiley: New York, 1989; Chapter 15.

26. *The Design and Analysis of Industrial Experiments*; Davies, O. L., Ed.; Hafner: New York, 1967; Chapter 7.

27. Blom, G. *Statistical Estimates and Transformed Beta Variables*; John Wiley: New York, 1958.

28. Box, G. E. P.; Cox, D. R., *J. Royal Stat. Soc. B*, **1964**, *26*, 211-252.

29. Mattu, M. J.; Small, G. W. *Anal. Chem.* **1995**, *67*, 2269-2278.

30. *Fed. Reg.* 54(12), 2332, 2955 (Jan. 19, 1989).

31. Hanst, P. L.; Stephens, E. R. *Spectrosc.*, 1989, *4(9)*, 33-38.

32. Hanst, P. L.; Wong, N. W.; Bragin, L. *Atmos. Environ.*, 1982, *16(5)*, 969-981.

33. Singh, H. B; Zimmerman, P. B. in *Gaseous Pollutants Characterization and Cycling*; Nriagu, J.O. Ed.; Wiley-Interscience: New York, 1992; Chapter 5.

34. Hanst, P. L. *Spectrosc.*, 1993, *8(9)*, 44.

35. Kroutil, R. T.; Ditillo, J. T.; Small, G. W. in *Computer-Enhanced Analytical Spectroscopy*; Meuzelaer, H., Ed.; Plenum: New York, 1990; Chapter 4.

36. Kroutil, R. T.; Combs, R. J.; Knapp R. B.; Small, G. W. *Proc. SPIE-Int. Soc. Opt. Eng.* 1993, *2089*, 24-31.

37. Combs, R. J.; Cathey, C. *Anal. Instrum.* 1992, *20*, 223-256.

38. Combs, R. J.; Kroutil, R. T.; Knapp, R. B. Unclassified Technical Report ERDEC-TR-298, Edgewood Research, Development, and Engineering Center, Aberdeen Proving Ground, MD, April, 1996.

39. Field, P. E.; Combs, R. J.; Knapp, R. B. Unclassified Technical Report ERDEC-TR-273,

Edgewood Research, Development, and Engineering Center, Aberdeen Proving Ground, MD, August, 1995.

40. Gmehling, J.; Onken, U.; Arlt, W. *Vapor-Liquid Equilibrium Data Collection*, Vol. 1, Part 8, DECHEMA: Frankfurt, Germany, 1984; p. 81.

41. Herzberg, G. *Molecular Spectra and Molecular Structure: II. Infrared and Raman Spectra of Polyatomic Molecules*; Van Nostrand Reinhold: New York, 1945; pp. 325-327.

42. Wu, T. *Phys. Rev.* 1934, *46*, 465-469.

43. Jansson, P. A. *Anal. Chem.* 1991, *63*, 357A.

44. Savitzky, A.; Golay, M. J. E. *Anal. Chem.* 1964, *36*, 1627-1639.

45. Steinier, J.; Termonia, Y.; Deltour, J. *Anal. Chem.* 1972, *44*, 1906-1909.

46. Low, M. J. D.; Clancy, F. K. *Environ. Sci. Technol.* **1967**, *1*, 73-74.

47. Griffiths, P. R. *Am. Lab.* **19:5**, *Mar*, 37-45.

48. Herget, W. F. *Am. Lab.* **1982**, *Dec*, 72-78.

49. Herget, W. F.; Brasher, J. D. *Appl. Opt.* **1979**, *18*, 3404-3420.

50. Polak, M. L.; Hall, J. L.; Herr, K. C. *Appl. Opt.* **1995**, *34*, 5406-5412.

51. Haus, R.; Schäfer, K.; Bautzer, W.; Heland, J.; Mosebach, H.; Bittner, H.; Eisenmann, T. *Appl. Opt.* **1994**, *33*, 5682-5689.

52. Haaland, D. M.; Easterling, R. G. *Appl. Spectrosc.* **1980**, *34*, 539-548.

53. Haaland, D. M.; Easterling, R. G.; Vopicka, D. A. *Appl. Spectrosc.* **1985**, *39*, 73-84.

54. Small, G. W.; Kroutil, R. T.; Ditillo, J. T.; Loerop, W. R. *Anal. Chem.* **1988**, *60*, 264-269.

55. Kroutil, R. T.; Housky, M.; Small, G. W. *Spectroscopy*, **1994**, *9*, 41-47.

56. Chan, S. H.; Lin, C. C.; Low, M. J. D. *Environ. Sci. Technol.* **1973**, *7*, 424-427.

57. Strang, C. R.; Levine, S. P.; Herget, W. F. *Am. Ind. Hyg. Assoc. J.* **1989**, *50(2)*, 70-77.

58. Strang, C. R.; Levine, S. P. *Am. Ind. Hyg. Assoc. J.* **1989**, *50(2)*, 78-84.

59. Ying, L. S.; Levine, S.P.; Strang, C. R.; Herget, W.F. *Am. Ind. Hyg. Assoc. J.* **1989**, *50(7)*, 354-359.

60. Ying, L. S.; Levine, S. P. *Anal. Chem.* **1989**, *61*, 677-683.

61. Mattu, M. J.; Small, G. W.; Arnold, M. A. *Appl. Spectrosc.* **1997**, *51*, 1369-1376.

62. Mattu, M. J.; Small, G. W.; Arnold, M. A. *Anal. Chem.* **1997**, *69*, 4695-4702.

63. Herget, W. F. *Appl. Opt.* **1982**, *21*, 635-641.

64. Martens, H.; Næs, T. *Multivariate Calibration*; Wiley: New York, 1989; Chapter 3.

65. Herget, W. F.; Brasher, J. D. *Opt. Eng.* **1980**, *19*, 508-514.